

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 07:56:07 ; Search time 3479.7 Seconds
(without alignments)
17070.103 Million cell updates/sec

Title: US-09-845-416-2_COPY_960_3000
Perfect score: 2041
Sequence: 1 tcttcacagcatttgaag.....acctcagcactctggaagac 2041

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 3813070

Minimum DB seq length: 0
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl: *
1: gb_ba: *
2: gb_htg: *
3: gb_in: *
4: gb_om: *
5: gb_ov: *
6: gb_pat: *
7: gb_ph: *
8: gb_pl: *
9: gb_pr: *
10: gb_ro: *
11: gb_sts: *
12: gb_sy: *
13: gb_un: *
14: gb_vi: *
15: em_ba: *
16: em_fun: *
17: em_hum: *
18: em_in: *
19: em_mu: *
20: em_om: *
21: em_or: *
22: em_ov: *
23: em_pat: *
24: em_ph: *
25: em_pl: *
26: em_ro: *
27: em_sts: *
28: em_un: *
29: em_vi: *
30: em_htg_hum: *
31: em_htg_inv: *
32: em_htg_other: *
33: em_htg_mus: *
34: em_htg_pln: *
35: em_htg_rod: *
36: em_htg_mam: *
37: em_htg_vrt: *
38: em_sy: *
39: em_htgo_hum: *
40: em_htgo_mus: *
41: em_htgo_other: *

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	851.6	41.7	4402	6	E30220	E30220 Shortened d
2	822.6	40.3	3275	10	MUSDYS	M18025 Mouse dyst
3	658.8	32.3	4075	6	E30221	E30221 Shortened d
4	654.6	32.1	4402	6	E30219	E30219 Shortened d
5	624.2	30.6	630	9	HSDMDF1	X06179 Human fetal
6	527.8	25.9	1966	9	HSDMDA1	X06178 Human adult
7	349.2	17.1	3747	6	E30218	E30218 Shortened d
8	344.6	16.9	1898	10	BC024140	BC024140 Mus muscu
9	344.4	16.9	3521	5	AF3339031	AF3339031 Danio rer
10	279.2	13.7	2654	5	FSCDYSTR0	M37645 Torpedo cal
11	273.4	13.4	500	9	HUMDYSTR15	L05649 Homo sapien
12	267.6	13.1	3161	10	MMGUTRPH	X83506 M.musculus
13	211.6	10.4	3499	9	HSU43519	U43519 Human dyst
14	210	10.3	3073	10	AF195788	AF195788 Rattus no
15	210	10.3	3140	10	AF195787	AF195787 Rattus no
16	198.2	9.7	1500	6	AX107969	AX107969 Sequence
17	198.2	9.7	1500	6	AX107970	AX107970 Sequence
18	193.2	9.5	645	9	AF213410	AF213410 Homo sapi
19	187	9.2	525	9	AF213411	AF213411 Homo sapi
20	180	8.8	454	9	AF213415	AF213415 Homo sapi
21	175.8	8.6	360	9	HUMDYSTR13	L05647 Homo sapien
22	161.2	7.9	618	9	AF213440	AF213440 Homo sapi
23	160.4	7.9	343	9	HUMDYSDMD	L04186 Homo sapien
24	152.8	7.5	352	9	AF213412	AF213412 Homo sapi
25	150	7.3	234	9	S60971	S60971 dystrophin
26	142.4	7.0	811	6	I07692	I07692 Sequence 15
27	131	6.4	375	9	HUMDMDF	M63075 Human Duch
28	125.6	6.2	456	9	AF213413	AF213413 Homo sapi
29	122	6.0	332	9	HUMDYSTR01	M86884 H.sapiens d
30	122	6.0	360	9	HUMDYSTR14	L05648 Homo sapien
31	120.4	5.9	989	9	HSDYSTX60	Z11860 H.sapiens g
32	115.2	5.6	726	9	AF213414	AF213414 Homo sapi
33	112.2	5.5	3650	3	SPAJ3356	AJ223356 Strongylo
34	111.2	5.4	1048	9	AF047502	AF047502 Homo sapi
35	97.6	4.8	400	10	S76214	S76214 DMD-Duchenn
36	81.4	4.0	2962	3	DMDYDTRO	X99757 D.melanogas
37	79.4	3.9	200	6	A63603	A63603 Sequence 5
38	78.6	3.9	200	6	A63602	A63602 Sequence 4
39	78.6	3.9	200	6	A63604	A63604 Sequence 6
40	63.6	3.1	111	5	AF137132	AF137132 Triakis s
41	63.6	3.1	120	5	AF137133	AF137133 Acipenser
42	63.6	3.1	238	6	AX114301	AX114301 Sequence
43	54.8	2.7	144	9	S60972	S60972 dystrophin
44	53.8	2.6	114	5	AF137135	AF137135 Osteoglos
45	52	2.5	117	5	AF137162	AF137162 Anguilla

ALIGNMENTS

RESULT 1
E30220
LOCUS E30220
DEFINITION Shortened dystrophin.
ACCESSION E30220
VERSION E30220.1 GI:13017027
KEYWORDS JP 1999318467-A/3.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 4402)
AUTHORS Sinichi,T.
TITLE Shortened dystrophin
JOURNAL Patent: JP 1999318467-A 3 24-NOV-1999;
SCIENCE & TECH AGENCY,NATIONAL CENTER OF NEUROLOGY AND PSYCHIATRY

4402 bp DNA linear PAT 18-JUN-2001

[illegible]

pr 28 09:28:32 2003

[illegible]

RESULT 3	linear	DNA	4075 bp	PAT 18-JUN-2001
E30221				
LOCUS				
DEFINITION				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
E30221				
Shortened dystrophin.				
E30221				
ACCESSION				
VERSION				
KEYWORDS				
SOURCE				
E30221.1				
GI:13017028				
JP 1999318467-A/4.				
unidentified.				

TITLE Shortened dystrophin
JOURNAL Patent: JP 199318467-A 1 24-NOV-1999;
SCIENCE & TECH AGENCY, NATIONAL CENTER OF NEUROLOGY AND PSYCHIATRY
COMMENT OS Unidentified
PN JP 199318467-A/1
PD 24-NOV-1999
PF 08-MAY-1998 JP 1998142134
PR
PI SINICHI TAKEDA
PC C12N15/09, A61K48/00, C12N15/00
CC Strandedness: Both;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..3747
/organism='Unidentified'.
FEATURES
source
Location/Qualifiers
1..3747
/organism='Unidentified'
/db_xref='taxon:32644' 923 t
BASE COUNT 1112 a 853 c 859 g 923 t
RIGIN
Query Match 17.1%; Score 349.2; DB 6; Length 3747;
Best Local Similarity 91.8%; Pred. No. 6.7e-74;
Matches 369; Conservative 0; Mismatches 33; Indels 0; Gaps 0;
Y 1 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAAGTCAATTTGGCAGTTTCATTGATGGAGAG 60
b 1155 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAAGTCAATTTGGCAGTTTCATTGATGGAGAG 1214
Y 61 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAGTATTATCGTGGCTTCTTC 120
b 1215 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAGTATTATCGTGGCTTCTTC 1274
Y 121 TGCTGAGGACACATTCGAAGCACAAGGAGAGATTTCTAATGATGTGGAGTGGTGAAGA 180
b 1275 TGCTGAGGACACATTCGAAGCACAAGGAGAGATTTCTAATGATGTGGAGTGGTGAAGA 1334
Y 181 CCAGTTTCATCTACTCATGAGGGGTACATGATGGATTTGACAGCCCATCAGGCGCGGTGG 240
b 1335 CCAGTTTCATCTACTCATGAGGGGTACATGATGGATTTGACAGCCCATCAGGCGCGGTGG 1394
Y 241 TAATATCTACAATTTGGGAAGTAAGCTGATTGGAACACAGGAAATATATCAGAAGATGAAGA 300
b 1395 TAATATCTACAATTTGGGAAGTAAGCTGATTGGAACACAGGAAATATATCAGAAGATGAAGA 1454
Y 301 AACTGAACTACAAGAGCAGATGAATCTCCTAAATTCAGATGGGAATGCCCTCAGGGTAGC 360
b 1455 AACTGAACTACAAGAGCAGATGAATCTCCTAAATTCAGATGGGAATGCCCTCAGGGTAGC 1514
Y 361 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGA 402
b 1515 CGTCGAGGACCGAGTCAGGCAGCTGCATGAAGCCCAACAGGGA 1556
RESULT 8
BC024140
DEFINITION Mus musculus, clone IMAGE:5149431, mRNA.
BC024140
VERSION BC024140.1 GI:18848250
KEYWORDS
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 1898)
Strausberg, R.
Direct Submission
Submitted (19-FEB-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK
COMMENT

NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Lothar Hennighausen Ph.D., Robin Humphreys
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Genome Sequence Centre,
BC Cancer Agency, Vancouver, BC, Canada
info@bcgsc.bc.ca
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,
Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,
Leticia Hsiao, Martin Krzywinski, Reta Kutsche, Oliver Lee, Soo
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven
Ness, Pawan Pandoh, Anna-Liisa Prabh, Parvaneh Saeedi, Jacqueline
Schein, Duane Smalilus, Michael Smith, Lorraine Spence, Jeff Stott,
Michael Thorne, Miranada Tsai, Natasja van den Bosch, Jill Vardy,
George Yang, Scott Zuyderduyn, Marco Marra.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 57 Row: p Column: 15
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis.

FEATURES
source

Location/Qualifiers
1..1898
/organism='Mus musculus'
/db_xref='taxon:10090'
/map='C57BL/6J'
/clone='IMAGE:5149431'
/tissue_type='Mammary tumor. WAP-TGF alpha model. 7 months
old, gross tissue.'
/clone_lib='NCI_CGAP_Mam5'
/lab_host='DH10B'
/note='Vector: pCMV-SPORT6'
BASE COUNT 582 a 404 c 540 g 372 t
ORIGIN

Query Match 16.9%; Score 344.6; DB 10; Length 1898;
Best Local Similarity 59.1%; Pred. No. 8.8e-73;
Matches 590; Conservative 0; Mismatches 409; Indels 0; Gaps 0;

QY 44 AGTTTCATGATGGAGAGTGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAGTA 103
Db 27 AGCACCGTCACTGAAGTGGACATGGATTTGGACAGCTACCAGATAGCGCTAGAGGAAGTG 86
QY 104 TTATCGTGGCTTCTTCTGCTGAGGACACATTCGCAAGCACAAGGAGAGATTTCTAATGAT 163
Db 87 CTGACGTGGTGTGTCGCGGAGGACACGTTCCAGGAGCAAGATGACATTTCTGATGAT 146
QY 164 GTGGAAGTGGTGAAGACACAGTTTCATCTCATGAGGGGTACATGATGGATTTGACAGCC 223
Db 147 GTCGAAGAAGTCAAGAGAGCAGTTTGTCTACCCATGAAACTTTATGATGGAGCTGACAGCA 206
QY 224 CATCAGGCGCGGTTGGTAATATTCTACAAATGGGAAGTAAGCTGATTGGAACAGGAAA 283
Db 207 CACCAGAGCAGCGTGGGAGCGCTCTGACAGGCTGGCAACAGCTGATGACACAAGGGACT 266
QY 284 TTATCAGAAGATGAAGAACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGATGG 343
Db 267 CTGTCAGAGGAGGAGGAGTTTGAGATCCAGGAACAGATGACCTTGTCTGAATGCAAGGTGG 326
QY 344 GAATGCCTCAGGGTAGCTAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGAT 403
Db 327 GAGGCGCTCCGGTGGAGAGCATGGAGAGGAGCTCCCGGCTGCACGACGCTCTGATGGAG 386
QY 404 CTCCAGAATCAGAAACTGAAGAGTGAATGACTGGCTAAACAAAACAGAAAGAAACA 463
Db 387 CTGCAGAAGAAACAGCTGCAGCAGCTCTCAAGCTGGCTGGCCCTCACAGAAGAGCGCAT 446
QY 464 AGGAAAATGGAGGAAGAGCGCTCTTGACCTGATCTTGAAGACCTAAACCGCCCAAGTACAA 523
Db 447 CAGAAGATGGAGAGCGCTCCCGCTGGGTGATGATGACCTGCCCTCCCTGCAGAAAGCTGCTTCAA 506

QY	524	CAACATAAGGTGCTTCAAGAAGATCTAGAACAAAGAACAAAGTCAAGGTCAATTCCTCACT	583
Db	507	GAACATAAAATTTGCAAAATGACCTTGAAGCTGAACAGGTGAAGTAAATTCCTTAAC	566
QY	584	CACATGGTGGTGTAGTTGATGAATCTAGTGGAGATCAGCAACTGCTGCTTGGGAAGAA	643
Db	567	CACATGGTGGTGTAGTTGGATGAACACAGTGGGAGAGTGGCCACAGCTCTTCTGGAAGAT	626
QY	644	CAACTTAAGGTATTGGGAGATCGATGGGCAAAACATCTGTAGATGGACAGAACCGCTGG	703
Db	627	CAGTTACAGAAACTGGGTGAGCGCTGGACAGCTGTATGCCGCTGGACTGAAGAACGTTGG	686
QY	704	GTTCTTTTACAGACATCCTTCTCAAAATGGCAACGCTTACTGAAGAACAGTGCCTTTT	763
Db	687	AACAGGTTGCAAGAAATCAGTATCTGTGGCAGGAAATTTGGAAGAGCAGTGTCTGTTG	746
QY	764	AGTGCATGGCTTTTCAGAAAAGAAAGATGCAGTGAACAAGATTCACACAACCTGGCTTAA	823
Db	747	GAGCTTGGCTCACCGAAAAGAGAGGCTTGAATAAGTTCAAACACAGCAACTTTAA	806
QY	824	GATCAAAATGAATGTTATCAAGTCTTCAAAAACCTGGCCGTTTAAAGCGGATCTAGAA	883
Db	807	GACCAGAAGGAACTAAGTGTCAAGTGTCCGGCTGTGGCTATATTGAAGGAACATGAA	866
QY	884	AAGAAAAGCAATCCATGGGCAAACTGTATTCACTCAAAACAGATCTTCTTCAACACTG	943
Db	867	ATGAAGAGCCAGACTCTGGATCAACTGAGTGAGATGGCCAGGATGGGCCAATTACTC	926
QY	944	AAGATAAGTCAAGTCAAGCAAGAGACGGAAGCATGGCTGGATAACTTGGCCGGTGG	1003
Db	927	ACTAATCCCAAGGCATCTAAGAAAGATGAACAGTGAACAGTCTAGGAGCTAACACAGATGG	986
QY	1004	GATAATTTAGTCAAAAACCTTGAAAGAGTACAGCACAG	1042
Db	987	GATTCTCTGGTTCAGAGACTCGAAGACTCTTCTCAACAG	1025

RESULT 9
AF339031
LOCUS
DEFINITION Danio rerio dystrophin (dmd) mRNA, partial cds, alternatively spliced.
ACCESSION AF339031
VERSION AF339031.1 GI:13699249
KEYWORDS
SOURCE Danio rerio.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 3521)
Bolanos-Jimenez, F., Borda, A., Behra, M., Strahle, U., Sahel, J. and Rendon, A.
Dystrophin and Dp71, two products of the DMD gene, show a different pattern of expression during embryonic development in zebrafish Mech. Dev. 102 (1-2), 239-241 (2001)
JOURNAL MEDLINE 21184125
PUBMED 11287201
REFERENCE 2 (bases 1 to 3521)
Bolanos-Jimenez, F., Rendon, A. and Strahle, U.
Direct Submission
Submitted (22-JAN-2001) Laboratoire de Physiopathologie Retinienne, EMI 99-18 INSERM-Universite Louis Pasteur, 1, Place de l'Hopital, Strasbourg 67091, France
Location/Qualifiers
1. .3521
/organism="Danio rerio"
/db_xref="taxon:7955"
/chromosome="1"
/map="between z5508 and z5058"
<1. .3521
/gene="dmd"

CDS

<1. .3521	/gene="dmd"	
/note="alternatively spliced"		
/codon_start=2		
/product="dystrophin"		
/protein_id="AAK38376.1"		
/db_xref="GI:13699250"		
/translation="RFCRHWEDSHAKLTARVLTQNMVKDSSDWLEARKRVEPLIKKA KVQMTESMNLAWANIKKRGADKLEAGLQHQHYLDLEKFLNLTAEATTANVL QDATFEGLEENPATVRLLEOWDLQAEIDAHEHYHSLDENGHRIVSSLEGTDNAV VLQKRLDDMGQRWHELCKNKMVSIRPYLDAGVDQWKHLHMSLOELLNLWLQKREELEKQ KPVGGDPTVHQQLLTHKAFRELGAKEPVIINGTLDNAKTLAEMPREGILKQRPQOKD VSPERVQNVGRILRKEVEDVTVRWKNLGAASVDWQOQLELALERLMELQADQDLDY KLQAESVKNWKPVGELLVDDLNHIDRVKAFQEEIAPIDQNVNHNQQLASTFRPSD IQLSPDNLSDIDDLNMRWLLQISIEEHLSTTAFKDLGPSONFLHASVESPLERSI SPNNVYYINHQTCTCDWDHPKMAELYQSLSLHLSLTAFKDLGPSONFLHASVESPLERSI LSMPAAACEAFEOHNLKQNEQMDIVQVINCILTSIYDRLEQOHSLSLVNPLCVDMCLNW LLNVYDTGRAGKINTLSFKTGIISLCAHLEDKYRFLREVASATGFCDDRRLGLLLH DAIQIPQLGEVASFGGSNIIEPSVRSFQFANNKPELEASVDLMMRLEPOSMVWLPV LHRVAAAEAKHQAKCNICKECPIIGFRYRSLKHFNYDQCSCFFSGRVAKHKMQYP MVEYCTPTSGEDVDVDFAKVLKNKFKRYFAKHPRMGYLPVQITILEGDNMETPVTLI NEFVDPHPASSPOLSHDDTHSRIEYASRDDEHLIIQHYCOSLNOGSPISQPSPAQ ILISMETEEKGELERLNLEQENRKLOAEYDRILKAHDKGLSPLEPQMLPVSPQ SPRAELIAEAKLLRQHKRLEARMQILEDHKNQLESQTLRLQLLEQTESKVNQTAL SSPTASPRSDTSLASLRVAASQTTETMGDELSSPTQDASTGLEVDIEQLNNSFPHS QGGGRLNP"		
3'UTR		
BASE COUNT	1005 a	892 g 756 t
ORIGIN		
Query Match		
Best Local Similarity	16.9%;	Score 344.4; DB 5; Length 3521;
Matches	592; Conservative	0; Mismatches 391; Indels

QY	1056	TGCAACAGTTCCTCCCTGGACCTGGAAAGTTCTTTCCTGGCTTACAGAACTGAAACAA	1115
Db	408	TGCAGCATTTACTACTTGGACCTGGAGAAATTCCTTAATGGCTAACGGAGGAGAACCA	467
QY	1116	CTGCCAATGTCTTACAGGATGCTACCCGTAAGGAAAGGCTCTCTAGAGACTCCAAAGGAG	1175
Db	468	CAGCAATGTCTTACAGATGCCACCTTTAAGAGGGACTTCTGSAATCTGCCCACAG	527
QY	1176	TAAAGAGCTGATGAAACAAATGGCAAGACCTCCAAAGGTGAAATTTGAAGCTCACACAGATG	1235
Db	528	TTTCGACATTTACTCGAGCAATGGCAGATCTCCAGGACAGATGATGCTCACCCGGAGA	587
QY	1236	TTTATCACAACTGGATGAAACAGACCCAAAATCTCTGAGATCCCTGGAAGGTTCCGATG	1295
Db	588	CGTACCATTCATTAGATGAAACAGGACATCGTATGTGTCTCTGGAGGAAACGGACA	647
QY	1296	ATGAGTCTCTCAACATTAGGTCCCATTTGGAAGCCAGTTCTGACCAAGTGAAGCGCTGC	1415
Db	648	ATGCTGTGTGCTACAAAACGGCTTGATGACATGGGGCAGCGCTGGCATGAGTTGTGCA	707
QY	1356	AAAAGTCTCTCAACATTAGGTCCCATTTGGAAGCCAGTTCTGACCAAGTGAAGCGCTGC	1415
Db	708	ACAAAGTTATGAGTATAAGGCCCTATCTAGATGCCGGCTTGATCAGTGGAAACACTTGC	767
QY	1416	ACCTTCTCTGCAGGAACCTTCTGGTGTGGCTACAGCTGAAAGATGATGAATTAAGCCGC	1475
Db	768	ACATGTCTTACAGAGCTGCTCAACTGGCTGACAGTGAAGGGAGAGCTGGAGAAGC	827
QY	1476	AGGCACCTATTGGAGCGGACTTTCCAGCAGTTCAGAGCAGAACCATGTACATAGGCCT	1535
Db	828	AGAAGCCAGTAGGGGCGACGTCGCCACCGTTCCACCAACACTCCTCACGCAACAGGCTT	887
QY	1536	TCAAGAGGAATTGAAAACCTAAAGAACCTGTAATCATGACTCTTGAGACTGTACGAA	1595
Db	888	TCAGAGAGAGCTGGGTGCCAAAGAACCTGTATCAATGGAACCTCTCGACAATCGGAAA	947
QY	1596	TATTTCTGACAGAGCAGCCTTTTGAAGGACTAGAGAACTCTACCAGGAGCCAGAGAGC	1655

CDS

QY	524	CAACATAAGGTGCTTCAAGAAGATCTAGAACAAAGAACAAAGTCAAGGTCAATTCCTCACT	583
Db	507	GAACATAAAATTTGCAAAATGACCTTGAAGCTGAACAGGTGAAGTAAATTCCTTAAC	566
QY	584	CACATGGTGGTGTAGTTGATGAATCTAGTGGAGATCAGCAACTGCTGCTTGGGAAGAA	643
Db	567	CACATGGTGGTGTAGTTGGATGAACACAGTGGGAGAGTGGCCACAGCTCTTCTGGAAGAT	626
QY	644	CAACTTAAGGTATTGGGAGATCGATGGGCAAAACATCTGTAGATGGACAGAACCGCTGG	703
Db	627	CAGTTACAGAAACTGGGTGAGCGCTGGACAGCTGTATGCCGCTGGACTGAAGAACGTTGG	686
QY	704	GTTCTTTTACAGACATCCTTCTCAAAATGGCAACGCTTACTGAAGAACAGTGCCTTTT	763
Db	687	AACAGGTTGCAAGAAATCAGTATCTGTGGCAGGAAATTTGGAAGAGCAGTGTCTGTTG	746
QY	764	AGTGCATGGCTTTTCAGAAAAGAAAGATGCAGTGAACAAGATTCACACAACCTGGCTTAA	823
Db	747	GAGCTTGGCTCACCGAAAAGAGAGGCTTGAATAAGTTCAAACACAGCAACTTTAA	806
QY	824	GATCAAAATGAATGTTATCAAGTCTTCAAAAACCTGGCCGTTTAAAGCGGATCTAGAA	883
Db	807	GACCAGAAGGAACTAAGTGTCAAGTGTCCGGCTGTGGCTATATTGAAGGAACATGAA	866
QY	884	AAGAAAAGCAATCCATGGGCAAACTGTATTCACTCAAAACAGATCTTCTTCAACACTG	943
Db	867	ATGAAGAGCCAGACTCTGGATCAACTGAGTGAGATGGCCAGGATGGGCCAATTACTC	926
QY	944	AAGATAAGTCAAGTCAAGCAAGAGACGGAAGCATGGCTGGATAACTTGGCCGGTGG	1003
Db	927	ACTAATCCCAAGGCATCTAAGAAAGATGAACAGTGAACAGTCTAGGAGCTAACACAGATGG	986
QY	1004	GATAATTTAGTCAAAAACCTTGAAAGAGTACAGCACAG	1042
Db	987	GATTCTCTGGTTCAGAGACTCGAAGACTCTTCTCAACAG	1025

/product="dystrophin"
/protein_id="AAC38002.1"
/db_xref="GI:397971"
/translation="SGEOWKRRLQISLDQFLTWMLNKDELRRQMPIGGDAPTVCQOND
VHRIFKRELKAKEPVMSALDVTVLEADPAIRAPESLLTPREKIP EENIQNVAKRI
RYAAEEVKVWEDKLSNRSVDWQKRIDEAKLRLLELQDSMDLNLKLRQAEAIKDTWOP
VGDLIDLSLDQHIIEKVKVFAEIAPMKENYTHMNDLASQFTPPDIQLSPYNLNQLEDL
NTRWKLQVSIIDELLYQLHEARHDFGTSQHF LSTSVQGPWERAI SPNKVPYYINHOT
QTCWDHPKMTLELLQSLADLNNVRFSAYRTAMKLRQLKALDL LSLPSACEAFDQH
NLKONDQLLDILEI INCLTSYDRLEQESHNVLPVLCVDMCLNWLNVYDTGRTGKI
RVLSFGSGIMSMCKAHLEDKYRYLFKQVASPTGFCDDORRLGLLHHEAIQTPRQLGEVA
SFGGSIENPSVSRCSQFQANNKPEIEAALFLDMRLEPQSLVWMPVLVRYAAETAQHO
AKCNICEKPIIGFRYRSLLKHFNVDVCQSCFFSGRTAGKHGMHYPMVEYCTPTSGED
VRDFAKVLKNFTKRYKFAKHPRMGYLPVQTVLEGDNLLETPTVLINFWPVDYEPASSP
QLSHDDTHSRTEHLYLSRAELMENRNGSYLNDISI SPNESIDDEHLLIQHYCQSLNQESP
LSQPSPAQILISEEERGELERILADLEDENRNQSEYEKIKQQDHKGLSPURSP
PEMPPI SPQSPDRAELIAEAKLLRHQKGRLEARMQILEDHNKQLSEQLHRLRLQLEQP
QAEVRVNGTSSVSPSTSSQRSDSSQPVVLHGVSQTS GILGEDDLLLSAPQSTSELED
VMEQLSSSFSSQRRDTPFGKKMKKEVPM"
c16 c16 617 t

PRASE COUNT		773 a	618 C	046 G	0
DRIGIN					
Query Match	13.7%;	Score 279.2;	DB 5;	Length 2654;	0;
Best Local Similarity	64.3%;	Pred. No. 6.6e-57;	Indels	0;	Gaps
Matches 419;	Conservative	0;	Mismatches 233;		
QY	1389	CCAGTCTTGACCAGTGGAAAGCGTCTGCACCTTCTCTGCAGGAACCTTCTGGTGTGGCTAC	1448		
Db	1	CAAGTGGTGACAGTGGAAAGCGGTGCAGATCTCCCTCCAGGACTTCCCTGACGTGGATGA	60		
QY	1449	AGCTGAAAGATGATGAATTAAGCCGCGCAGGCACCTATTGGAGCGCACTTCCAGCAGCTTC	1508		
Db	61	ATTGGAAGAACGATGAACATAAGGAGGCAAAATGCCCATTTGGTGTGATGCTCCACACAGTCT	120		
QY	1509	AGAAGCAGAACGATGACATAGGGCCTTCAAGAGGGAATTTGAAAACCTAAAGAAACCTGTAA	1568		
Db	121	GTACGACAACGATGTTACCGAATTTTCAAGCGAGAGTTGAAGGCAAGGAACCAAGTGG	180		
QY	1569	TCATGAGTACTCTTTGAGACTGTACGAATATTTCTGCACAGCAGCCTTTGGAGGAGCTAG	1628		
Db	181	TCATGAGCGCTTTGGACACAGTGCAATTTGTTCTTGGCTGATCCAGCAATCAGAGCTCCTG	240		
QY	1629	AGAAACTCTACAGGAGCCAGAGAGCTGCCTCTGAGGAGAGAGGCCAGAAATGTCACTC	1688		
Db	241	AAAGTCTTTAACTGGACCAAGAGAGAAAATACCTGAAGAGAACTCCAAAATGTTGCGA	300		
QY	1689	GGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTGAACCTGCAC	1748		
Db	301	AACGCATTCGGAAGTATGCCGAGGAGGTGAAAGTGGATGGGATAAGCTGAGCAATCGCT	360		
QY	1749	CCGCTGACTGGCAGAGAAAATAGATGAGACCCCTTGAAGAGACTCCAGGAACCTTCAAGAG	1808		
Db	361	CTGTTGATTGGCAGAAGCGTATAGATGAGCCTTGAAGAGACTACTGGAATTCGAAGATT	420		
QY	1809	CCACGGATGAGCTGGACCTCAAGCTGGCCAGCTGAGGTGATCAAGGGATCCTGGCAGC	1868		
Db	421	CAATGGATGAATTGAACCTCAAATTTGAGACAGGCTGAAGCTATCAAAGATACATGGAAC	480		
QY	1869	CCGTGGCGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGAGAAAAGTCAAGGCACTTC	1928		
Db	481	CTGTCGGGGATCTACTAATAGACTCATTTGCAGGATCACATTGAAAAGTCAAGGTTTTC	540		
QY	1929	GAGGAGAAATTCGCCCTCTTGAAGAGAACTGAGCCACCTCAATGACCTTGCTCGCCAGC	1988		
Db	541	GAGCAGAAATTCCTCCCATGAAGGAAAATGTGACTCACATGAATGATCTCGCTTCCCAGT	600		
QY	1989	TTACCACTTTGGGCATTACGCTCTCACCGTATAAECTCAGCACTCTTGGAGA	2040		
Db	601	TCACACCACTGATATCCAAATTATCCCCGTACAATCTAAACCAAGTTGGAGGA	652		

RESULT 11
HUMDYSTR15

```

RESULT 10
FSCDYSTRO
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

FSCDYSTRO
Torpedo californica dystrophin mRNA, 3' end.
M37645
M37645.1 GI:397970
dystrophin.
Torpedo californica, cDNA to mRNA.
Torpedo californica
Torpedo californica
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
Elasmobranchii; Squalea; Hypnosqualea; Pristiogaster; Batoidae;
Torpediniformes; Torpedinoidei; Torpedinidae; Torpedo.
1 (bases 1 to 2654)
yeaton,J.E.; Lin,H.; Dyer,S.M. and Burden,S.J.
Dystrophin is a component of the subsynaptic membrane
J. Cell Biol. 115 (4), 1069-1076 (1991)
92064638
1720119
2 (bases 1 to 2654)
Ravin,A.J.; Dyer,S.M., Yeaton,J.E. and Burden,S.J.
Multiple dystrophin isoforms are associated with the postsynaptic
membrane of Torpedo electric organ
J. Physiol. (Paris) 85 (3), 131-133 (1991)
92291902
1818109
On Sep 8, 1993 this sequence version replaced gi:213232.
Draft entry and computer-readable sequence for [Unpublished (1990)
kindly submitted
by S.J.Burden, 06-AUG-1990.
Biology Dept, 16-820
MIT
Cambridge, MA 02139.
Location/Qualifiers
1..2654
/organism="Torpedo californica"
/db_xref="taxon:7787"
1..2654
/gene="dystrophin"
<1..2654
/gene="dystrophin"
/codon_start=3

FEATURES
source
gene
CDS

```


BASE COUNT

BASE COUNT

704 C	802 g	688 L
-------	-------	-------

us-09-845-416-2_copy_960_3000.rge

Mon Apr 28 09:28:32 2003

KEYWORDS									
SOURCE									
ORGANISM									
Homo sapiens.									
Homo sapiens									
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;									
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.									
1 (bases 263 to 3189)									
Robertson, R.G., Freeman, F.C., Kendall, E., Vetric, D.L., Dixon, A.K.,									
Shaw-Smith, C., Bone, Q. and Bobrow, M.									
Characterization of DRP2, a novel human dystrophin homologue									
Nat. Genet. 13 (2), 223-226 (1996)									
TITLE									
JOURNAL									
MEDLINE									
PUBMED									
2 (bases 1 to 3499)									
Robertson, R.G.									
Direct Substitution									
Submitted (16-DEC-1995) Roland G. Roberts, Medical Genetics,									
Addenbrooke's Hospital, 3rd Floor, Lab Block, Addenbrooke's									
Hospital, Hills Road, Cambridge, Cambs CB2 2Qq, UK									
Location/Qualifiers									
1. 3499									
/organism="Homo sapiens"									
/db_xref="taxon:9606"									
/chromosome="X"									
/map="Xq22"									
1. 3499									
/gene="DRP2"									
<1. 262									
/gene="DRP2"									
328. 3192									
/gene="DRP2"									
/function="membrane cytoskeleton"									
/codon_start=1									
/product="dystrophin-related protein 2"									
/protein_id="AAC50538.1"									
/db_xref="GI:1353782"									
/translation="MVMQGGPYTLPRCHDQAAADQFHSSSLRSTCPHPQVRAAVTSP									
APPQDAGVPCLSKLLNGSVGASGPPEPPAMNLCWNEIKKSHNLRALEAFSDHSG									
KLQPLQELIDWLSQKDELSAQLPLQGDVALVQOEKETHAFAEVEVSKRPIYISVL									
ESQAFLSQHPFEELEPHSEKSTSPKQRIQNLRFVWKQATVASELWEKLTARCDV									
QHRHIERLEQLLEIQAMEELSTLSOAEVVRATWEP IGDLFIDSLPEHIQAIKLFK									
EEDSPMKDGVKLVNDLAHOLAISDVHLSMENSQALQINVRWKQLQASVDERLKLQD									
AHRDFGPGSQHFLSSSVQVWERATSPNKVPY ITHQAOCTCDWHPKMTVELYHCLTA									
LYERLEERGIILVNPCLVDMSLNWLNVFDSGRSGKMRALSKFTGIACLCGTEVEK									
LQVLFQVANSQSQDQDRLHGLVLLHEAICQVPRQLGEVAAFGGSNVEPSVRSRFSFG									
KPVIEASQFLVNLFPQSMVWLPVLRHTVIAEQVHQKTKSCQKQKPIKGFYRSK									
QFVNDICQTCFLTGASQKGLHYPIMEYITPTSSNMDFATLKNKFRSKHYFSL									
HPQGVLPVQSVLEADYSETPASSPMWPHADHSRIEFAELAESQNSQSFNRLQG									
SPDSDIDQYLLRHSSPTDREPAGFQQAPCSVATESKELQKILAHLEDENRILQG									
ELRLKQWHEEAEEAPSLADGSTEATDHRNEELAEARILRQHKSRLETRMQILEDH									
NKQLESQRLRLRELLLPPTESDGSAGSSSLASSPQSGSHPREKQGTTPDTEAAD									
DVGSKSQDVSLCLEIDIMEKLRAFPSPVSRSSDVTANTLLAS"									
3193. ->3499									
/gene="DRP2"									
3'UTR									
BASE COUNT									
ORIGIN									
Query Match									
Best Local Similarity									
Matches									
10.4%; Score 211.6; DB 9; Length 3499;									
Pred. No. 1.7e-40;									
0; Mismatches 314; Indels									
0; Gaps									
Query									
1327 CATGAACCTCAAGTGGAGTGAACCTCGGAAAAGTCTCTCAACATTAGTCCCATTTGGA									
1386									
Db									
552 CATGAATCTGTGTGAATGAATAAAGAGAGTCTCAACCTCCGCGCTCGCTAGA									
611									
QY									
1387 AGCCAGTCTTGACCACTGGAAGCTGTGACCTTCTCTGAGGAACCTTCTGTGTGGCT									
1446									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1387 AGCCAGTCTTGACCACTGGAAGCTGTGACCTTCTCTGAGGAACCTTCTGTGTGGCT									
1446									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									
1447 ACAGCTGAAAGATGATGAATTAAAGCCGCGGAGGACCTATTGGAGGCGGAGTTCACGACGT									
1506									
Db									
612 GGCCTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTTCAAGAGATTATTGACTGGCT									
671									
QY									

Mon Apr 28 09:28:32 2003

1927 TCGAGGAGAAATTGGCGCTCTGGAAGAGAACGTCAGCCACGTCATGACCTTGCTCGCCA 1986
1033 CAAAGAGAATTTCTCTCTGTGAAGACGGGGTGAAGTTAGTGAATGACCTGGCCACCA 1092
1987 GCTTACCACTTTGGGCATTCAGCTCTACCCGTATAAACCTCAGCACTCTGGAAGA 2040
1993 GCTTGCTATTCTGATGTGCACTTGTCATGAGGAGAAATTCAGGGCTCTGGAACA 1146

RESULT 15
AF195787 3140 bp mRNA linear ROD 02-MAY-2001
LOCUS Rattus norvegicus dystrophin-related protein 2 A-form splice
DEFINITION variant (Drp2) mRNA, complete cds.

AF195787.1 GI:11066164

Rattus norvegicus.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

1 (bases 1 to 3140)
Roberts,R.G. and Sheng,M.
Association of dystrophin-related protein 2 (Drp2) with
postsynaptic densities in rat brain

Mol. Cell. Neurosci. 16 (5), 674-685 (2000)
20538711
11083927
2 (bases 1 to 3140)
Roberts,R.G. and Sheng,M.

Direct Submission
Submitted (18-OCT-1999) Division of Medical and Molecular Genetics,
GKT Medical School, 8th Floor, Guy's Hospital, Guy's Tower, London
SE1 9RT, UK

Location/Qualifiers
1. .3140
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"

1. .3140
/gene="Drp2"
267. .3140
/gene="Drp2"
/note="Drp2; membrane-associated cytoskeletal protein;
contains two spectrin repeats; WW domain; ZZ domain"
/codon_start=1
/product="dystrophin-related protein 2 A-form splice
variant"
/protein_id="AAG28484.1"
/db_xref="GI:11066165"
/translation="MPLVMQGPYTLPRCHEWHAADRFHSSSLRNTCPQPVRAAV
TIPAPPWDGADPCLSKLLNGSVGAVGLEPSAMNLCWNEIKKSKHNLRLAEAFSD
HSGKLQPLQETIDWLSQKDELSAQPLQGDVALVQEKETHAAFMEEVKSQGPYIY
SVLESAQAFLSQHPFEELESHSESKDTSRQIQLSRFVWKQATVASELWEKLTAR
CVDQHRHIEHTLEHLEIQGMEELSSLTQAEGVRATWEPIGDLFIDSLPEHIQAIK
LFKEEFPVKDGVKLVNDLAHQLAISDVHLSMENSRALEQINVRWKQLQSVAEERLQ
LQDAHRDFGPGSQHFLSTSVQVWERAI SPNVYPYINHQAQTTCWDHPKMTELYQTL
ADLNNIKFSAYRTAMKLRVQKALRLDLVLTALTLEIFNEHDLQASEHVMVDVVEIHC
LTALYERLEEEERGLVNVPLCVDMSLNLNLFVDSRSGKMRALSFKTGIAACLCQTEV
KEKLQYLFVSQVANSQKCDQRLGALLHEAIQVPRQGEVAFAFGGSNVEPSVRSFRR
STGKPVIEASQFLEWNLFPQSMVWLAVLHRTVVAEQVKHQTCKSICRCQCPKGFYR
SLKQFNVDICQTCFLTGRASKNKLHYPIMEYTPPTSSNMRFATTLKNKFRSKQY
FSKHPQRYLPVQSVLESDCSETPASSPMLPHADTHSRHIEFASRLAEMESQNCSEFN
DLSLSPDSDIDEDQYLLRHSSPTDREPAPGQAPCPMATESKGELEKILAHLEDENRI
LQGLRLKQWHEEAAPTAEAEATPDHNEELAEARILRQHKSRLETRMQIL
EDHNKQLESQRLRELLOPPTESDNGSAGSSLPQSESGSHPREKGGQTTPTDTE
AADDVGSKSQDVSHSLEDIMEKLRHAFPSVRSDDVTANTLLAS"

BASE COUNT 790 a 864 c 794 g 690 t 2 others
ORIGIN
Query Match 10.3%; Score 210; DB 10; Length 3140;
Best Local Similarity 55.9%; Pred. No. 4.2e-40;

Matches 399; Conservative 0; Mismatches 315; Indels 0; Gaps 0;
QY 1327 CATGAACCTCAAGTGGAGTGAACCTTCGGAAAAAGTCTCTCAACATTAGGTCCCATTTGGA 1386
Db 500 CATGAATCTGTGTGAATGAAATAAAAAAGAGTCTCACAACCTCCGCGCTCGCCTAGA 559
QY 1387 AGCCAGTTCTGACCAGTGAAGCGCTCTGACCTTTCTCTGCAGGAACCTCTGGTGTGGCT 1446
Db 560 GGCCTTCTCAGACCACAGTGAAGAACTTCAGCTCCCACTTCAAGAGATTTAGTACTGGCT 619
QY 1447 ACAGCTGAAAGATGATGAATTAAGCCGGCAGGACCTATTTGGAGGCGACTTTCAGCAGT 1506
Db 620 CAGCCAAAAGGATGAAGAAATGTGAGCTCAGCTGCCCTTGCAGGGGATGTGGCCCTGGT 679
QY 1507 TCAGAACGAGAACGATGTACATAGGGCCTTCAAGAGGGAATTTGAAACTAAAGAACCTGT 1566
Db 680 ACAACAGGAGAGGAGACACATGCTGCCTTTATGGAAGAGTGAAGTCTAAAGGCCCTTA 739
QY 1567 AATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTTGGAGGACT 1626
Db 740 TATCTACTGTGTGCTGGAATCGGCACAGGCTTTCCTGTCTCAGCACCCTTGTGAGGAAT 799
QY 1627 AGAGAACTCTACAGGAGGCCAGAGAGTGCCTCCTGAGGAGAGAGCCAGCAATGTAC 1686
Db 800 AGAGGAGTCTCATTTGAGAGCAAGATACTCCCAAGACAGCGGATTCAAAAACCTCAG 859
QY 1687 TCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTTGAACCTGCA 1746
Db 860 TCGCTTTGTGTGAAGCAGGCAACAGTGGCCAGTGAACCTGAGGAGGGAGTCCGAGCCACATGGGA 1039
QY 1747 CTCCGCTGACTGGCAGAGAAAAAATAGATGAGACCTTGAAGAGACTCCAGGAACCTCAAGA 1806
Db 920 CTGTGTGGATCAGCATCGCCACATTTGAGACACACCTGGACATCTATTGGAGATCCAAG 979
QY 1807 GGCACGGATGAGCTGGACCTCAAGCTGCGCAAGCTGAGTGATCAAGGGATCCTGGCA 1866
Db 980 GGCAATGAGGAACCTGAGCAGTACTTTGACCCCAAGCAGAGGGAGTCCGAGCCACATGGGA 1039
QY 1867 GCCCGTGGCGGATCTCCTCATTTGACTCTCTCCCAAGATCACCTCGAGAAAGTCAAGGCACT 1926
Db 1040 GCCCATAGGAGATCTCTTTATCGATTCCCTCCAGAGCATATCCCAAGCCATCAAGTTATT 1099
QY 1927 TCGAGGAGAAATTCGCCCTCTGAAAGAGAACCTGAGCCACGTCAATGACCTTGTCTGCCA 1986
Db 1100 CAAAGAAGAAATTCCTCTCTGTGAAAGACGGGGTGAAGTTAGTGAATGACCTGGCCACCA 1159
QY 1987 GCTTACCACCTTTGGGCATTCAGCTCTCAACCTTACCTCAGCAGCTCTGGAAGA 2040
Db 1160 GCTTGCTATTCTGATGTGCACTTTGTCAATGGAGAAATTCAGGGGCTCTTGAACA 1213

Search completed: April 25, 2003, 14:25:53
Job time : 3532.87 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 07:25:57 ; Search time 284.596 Seconds
(without alignments)
16150.348 Million cell updates/sec

Title: US-09-845-416-2_COPY_960_3000

Perfect score: 2041

Sequence: 1 tccttcacagcatttgaag.....acctcagcactctggaagac 2041

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4313688

Minimum DB seq length: 0

Maximum DB seq length: 5000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_101002.*

- 1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
- 23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
- 24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2041	100.0	4182	24	AAD37230 Human dystrophin m
2	1519.8	74.5	3999	24	AAD37234 Human dystrophin m
3	1519.8	74.5	4966	24	AAD37256 Adeno-associated v
4	1519.8	74.5	4990	24	AAD37262 Adeno-associated v
5	1383	67.8	3858	24	AAD37237 Human dystrophin m
6	1383	67.8	4825	24	AAD37257 Adeno-associated v
7	1383	67.8	4848	24	AAD37263 Adeno-associated v
8	1054	51.6	4414	24	AAD37260 Adeno-associated v
9	1042	51.1	3446	24	AAD37242 Human dystrophin m

10	1037.8	50.8	1991	24	AAD37231 Human dystrophin N
11	999	48.9	2169	24	AAD37232 Human dystrophin r
12	999	48.9	3531	24	AAD37238 Human dystrophin m
13	999	48.9	4498	24	AAD37258 Adeno-associated v
14	851.6	41.7	4402	21	AAZ48568 A rod shortened dy
15	811.4	39.8	3275	10	AAN97129 Partial sequence o
16	727	35.6	3510	24	AAD37240 Human dystrophin m
17	727	35.6	4476	24	AAD37259 Adeno-associated v
18	717	35.1	1667	24	AAD37235 Human dystrophin N
19	658.8	32.3	4075	21	AAZ48569 A rod shortened dy
20	654.6	32.1	4402	21	AAZ48567 A rod shortened dy
21	652	31.9	1821	24	AAD37241 Human dystrophin r
22	390	19.1	1340	24	AAD37239 Human dystrophin N
23	387	19.0	387	24	ABK81990 Human dystrophin s
24	349.2	17.1	3747	21	AAZ48566 A rod shortened dy
25	348	17.1	348	24	ABK81989 Human dystrophin s
26	333	16.3	333	24	ABK81967 Human dystrophin s
27	327	16.0	327	24	ABK81966 Human dystrophin s
28	322.4	15.8	333	24	ABK81968 Human dystrophin s
29	265	13.0	1434	24	AAD37243 Human dystrophin r
30	261	12.8	324	24	ABK81991 Human dystrophin s
31	198.2	9.7	1500	22	AAF84672 Nucleotide sequenc
32	190.2	9.3	256	22	ABA69736 Human foetal liver
33	190.2	9.3	256	22	ABA36636 Probe #15102 for g
34	190.2	9.3	256	22	AAK17925 Human brain expres
35	190.2	9.3	256	22	AAK43799 Human bone marrow
36	190.2	9.3	256	22	AAI24574 Probe #14507 for g
37	190.2	9.3	256	22	AAI49820 Probe #18506 used
38	190.2	9.3	256	24	ABS18027 Human genome-deriv
39	180	8.8	476	24	ABS08442 Human genome-deriv
40	177	8.7	466	22	ABA58628 Human foetal liver
41	177	8.7	466	22	ABA27626 Probe #6092 for ge
42	177	8.7	466	22	AAK06756 Human brain expres
43	177	8.7	466	22	AAK32467 Human bone marrow
44	177	8.7	466	22	AAI16084 Probe #6017 for ge
45	177	8.7	466	22	AAI38301 Probe #6987 used t

ALIGNMENTS

RESULT 1

AAD37230	ID	AAD37230 standard; DNA; 4182 BP.
XX	XX	
AC	AC	AAD37230;
XX	XX	
DT	DT	21-AUG-2002 (first entry)
XX	XX	
DE	DE	Human dystrophin minigene delta4173.
XX	XX	
KW	KW	Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW	KW	adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW	KW	Becker muscular dystrophy; ds.
XX	XX	
OS	OS	Homo sapiens.
XX	XX	
PN	PN	WO200183695-A2.
XX	XX	
PD	PD	08-NOV-2001.
XX	XX	
PF	PF	27-APR-2001; 2001WO-US13677.
XX	XX	
PR	PR	28-APR-2000; 2000US-200777P.
XX	XX	
PA	PA	(XIAO/) XIAO X.
XX	XX	
PI	PI	Xiao X;
XX	XX	
DR	DR	WPI; 2002-049342/06.
XX	XX	
PT	PT	New dystrophin minigene for treating Duchenne or Becker muscular dystrophy comprises an N-terminal domain or modified N-terminal domain,

rod repeats, H1 and H4 domains and a cysteine rich domain of a dystrophin gene -

Example 1; Page 43-44; 71pp; English.

The present invention relates to an isolated nucleotide sequence encoding a dystrophin minigene. The minigene comprises N-terminal or modified N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control element. The dystrophin minigene in operable linkage with an expression control element, in a recombinant adeno-associated virus or retrovirus is useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is human dystrophin minigene delta4173 containing nucleotides 1-1992 (N-terminus, hinge H1 and rods R1, R2 and R3), 8059-10227 (rods R22, R23 and R24, hinge H4 and CR domain) and 11047-11058 (dystrophin last 3 amino acids).

Sequence 4182 BP; 1309 A; 927 C; 970 G; 976 T; 0 other;

Query Match	100.0%;	Score 2041;	DB 24;	Length 4182;	
Best Local Similarity	100.0%;	Pred. No. 0;			
Matches 2041;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
QY	1	TCCTTACACGATTTGGAAGCTCCTGAAGACAACTGATTTGGCAGTTTCATTGATGGAGAG	60		
Db	960	TCCTTACACGATTTGGAAGCTCCTGAAGACAACTGATTTGGCAGTTTCATTGATGGAGAG	1019		
QY	61	TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGATGATTTATCGTGGCTTCTTC	120		
Db	1020	TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGATGATTTATCGTGGCTTCTTC	1079		
QY	121	TGCTGAGGACACATTCGAAGCACAAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA	180		
Db	1080	TGCTGAGGACACATTCGAAGCACAAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA	1139		
QY	181	CCAGTTTCATCTACTGAGGGGTACATGATGGATTTTGACAGCCCATCAGGGCCGGGTGG	240		
Db	1140	CCAGTTTCATCTACTGAGGGGTACATGATGGATTTTGACAGCCCATCAGGGCCGGGTGG	1199		
QY	241	TAATATCTACAATTTGGGAAGTAAAGCTGATTTGGAACAGGAAATTTATCAGAAGATGAAGA	300		
Db	1200	TAATATCTACAATTTGGGAAGTAAAGCTGATTTGGAACAGGAAATTTATCAGAAGATGAAGA	1259		
QY	301	AACCTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGATGGGAATGCCATCAGGGTAGC	360		
Db	1260	AACCTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGATGGGAATGCCATCAGGGTAGC	1319		
QY	361	TAGCATGGAAGAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGATCAGAAACT	420		
Db	1320	TAGCATGGAAGAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGATCAGAAACT	1379		
QY	421	GAAAGAGTTGAATGACTGGCTAACAACAAACAGAAAGAAACAAAGGAAATGGAGGAAGA	480		
Db	1380	GAAAGAGTTGAATGACTGGCTAACAACAAACAGAAAGAAACAAAGGAAATGGAGGAAGA	1439		
QY	481	GCCTCTTGGACCTGATCTTGAAGACCTAAACCGCCCAAGTACAACAAACATAAGGTGCTTCA	540		
Db	1440	GCCTCTTGGACCTGATCTTGAAGACCTAAACCGCCCAAGTACAACAAACATAAGGTGCTTCA	1499		
QY	541	AGAAGATCTAGAACAAACAAAGTCAAGGTCAATTTCTCACTCACATGGTGGTGGTAGT	600		
Db	1500	AGAAGATCTAGAACAAACAAAGTCAAGGTCAATTTCTCACTCACATGGTGGTGGTAGT	1559		
QY	601	TGATGAATCTAGTGGAGTACCGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGG	660		
Db	1560	TGATGAATCTAGTGGAGTACCGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGG	1619		
QY	661	AGATCGATGGGCAACATCTGTAGATGGACAGAAGACCGCTGGGTCTTTTACAAGACAT	720		
Db	1620	AGATCGATGGGCAACATCTGTAGATGGACAGAAGACCGCTGGGTCTTTTACAAGACAT	1679		

QY	721	CCTTCTCAATGGCAACGCTTACTGAAGAACACAGTGCCTTTTGTAGTGCATGGCTTTCAGA	780
Db	1680	CCTTCTCAATGGCAACGCTTACTGAAGAACACAGTGCCTTTTGTAGTGCATGGCTTTCAGA	1739
QY	781	AAAAGAAGATGCAAGTGAACAAAGATTCACACAACTGGCTTTAAAGATCAAAATGAAATGTT	840
Db	1740	AAAAGAAGATGCAAGTGAACAAAGATTCACACAACTGGCTTTAAAGATCAAAATGAAATGTT	1799
QY	841	ATCAAGTCTTCAAAAACCTGGCCGTTTAAAGCGGATCTAGAAAAAGAAAAAGCAATCCAT	900
Db	1800	ATCAAGTCTTCAAAAACCTGGCCGTTTAAAGCGGATCTAGAAAAAGAAAAAGCAATCCAT	1859
QY	901	GGCAAACTGTATTCACCTCAAAACAAAGATCTTCTTCAACACACTGAAGAATAAGTCAGTGAC	960
Db	1860	GGCAAACTGTATTCACCTCAAAACAAAGATCTTCTTCAACACACTGAAGAATAAGTCAGTGAC	1919
QY	961	CCAGAAAGACGGAAGCATGGCTGGATAACTTTGCCCGGTGTTGGGATAATTTAGTCCAAAA	1020
Db	1920	CCAGAAAGACGGAAGCATGGCTGGATAACTTTGCCCGGTGTTGGGATAATTTAGTCCAAAA	1979
QY	1021	ACTTGAAGAAAGATACAGCACAGACTCATAGATTACTGCAACAGTTTCCCTCCCTGGACCTTGA	1080
Db	1980	ACTTGAAGAAAGATACAGCACAGACTCATAGATTACTGCAACAGTTTCCCTCCCTGGACCTTGA	2039
QY	1081	AAAGTTTCTTCTGCTGGCTTACAGAACTCCAAAGGAGTGAACAACTGCTTACAACTGCTTAC	1140
Db	2040	AAAGTTTCTTCTGCTGGCTTACAGAACTCCAAAGGAGTGAACAACTGCTTACAACTGCTTAC	2099
QY	1141	CCGTAAGGAAAGGCTCCTAGAAAGACTCCAAAGGAGTGAACAACTGCTTACAACTGCTTAC	1200
Db	2100	CCGTAAGGAAAGGCTCCTAGAAAGACTCCAAAGGAGTGAACAACTGCTTACAACTGCTTAC	2159
QY	1201	AGACCTCCAAGTGAATTTGAAGCTCACACAGATGTTTATCAACACCTGGATGAAACACAG	1260
Db	2160	AGACCTCCAAGTGAATTTGAAGCTCACACAGATGTTTATCAACACCTGGATGAAACACAG	2219
QY	1261	CAAAAAATCCTGAGATCCCTGGAAGGTTCCGATGATGCAGTCTTCAACAACTGCTTACAA	1320
Db	2220	CAAAAAATCCTGAGATCCCTGGAAGGTTCCGATGATGCAGTCTTCAACAACTGCTTACAA	2279
QY	1321	GGATAACATGAACCTTCAAGTGGAGTGAACCTCGGAAAAAGTCTCAACATTAGTGTCCCA	1380
Db	2280	GGATAACATGAACCTTCAAGTGGAGTGAACCTCGGAAAAAGTCTCAACATTAGTGTCCCA	2339
QY	1381	TTTGGAAAGCCAGTTCTGACCACTGGAAGGCTGTCACCTTTCTCTGCAGGAACTTCTGGT	1440
Db	2340	TTTGGAAAGCCAGTTCTGACCACTGGAAGGCTGTCACCTTTCTCTGCAGGAACTTCTGGT	2399
QY	1441	GTGGCTACAGCTGAAAGATGATGAATTAAGCCGCGCAGGCACCTATTGGAGGCGACTTTC	1500
Db	2400	GTGGCTACAGCTGAAAGATGATGAATTAAGCCGCGCAGGCACCTATTGGAGGCGACTTTC	2459
QY	1501	AGCAGTTTCAAGACAGAACCGATGTACATAGGSCCTTCAAGAGGGAATTTGAAAACTAAAGA	1560
Db	2460	AGCAGTTTCAAGACAGAACCGATGTACATAGGSCCTTCAAGAGGGAATTTGAAAACTAAAGA	2519
QY	1561	ACCTGTAATCATGAGTACTCTTGAGACTGTACGATATTTCTGACAGAGCAGCCTTTTGA	1620
Db	2520	ACCTGTAATCATGAGTACTCTTGAGACTGTACGATATTTCTGACAGAGCAGCCTTTTGA	2579
QY	1621	AGGACTAGAGAACTCTTACAGGAGCCCGCAGAGAGTGCCTCTGAGGAGAGAGCCAGAA	1680
Db	2580	AGGACTAGAGAACTCTTACAGGAGCCCGCAGAGAGTGCCTCTGAGGAGAGAGCCAGAA	2639
QY	1681	TGTCACCTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTGAA	1740
Db	2640	TGTCACCTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTGAA	2699
QY	1741	CCTGCACCTCCGCTGACTGGCAGAGAAAAAATAGATGAGACCTTGAAGAGACTCCAGGAAC	1800
Db	2700	CCTGCACCTCCGCTGACTGGCAGAGAAAAAATAGATGAGACCTTGAAGAGACTCCAGGAAC	2759
QY	1801	TCAAGAGGCCACCGATGAGCTGGACCTCAAGCTGCGCAAGCTGAGGTGATCAAGGGATC	1860

THE

Query Match		74.5%	Score 1519.8;	DB 24;	Length 3999;
Best Local Similarity		87.5%	Pred. No. 0;		
Matches 1786; Conservative		0;	Mismatches	72;	Indels 183; Gaps
QY	1	TCCTTCACAGCATTTGGAAAGCTCCTGAAGACAACTGATTTGGCAGTTCAATTCGTGGCTTCTTTC			
Db	960	TCCTTCACAGCATTTGGAAAGCTCCTGAAGACAACTGATTTGGCAGTTCAATTCGTGGCTTCTTTC			
QY	61	TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAGAAAGTATTCATCGTGGCTTCTTTC			
Db	1020	TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAGAAAGTATTCATCGTGGCTTCTTTC			
QY	121	TGCTGAGGACACATTTGCAAGCACAAGGAGAGATTTCTTAATGATGTGGAAGTGGTGAAGA			
Db	1080	TGCTGAGGACACATTTGCAAGCACAAGGAGAGATTTCTTAATGATGTGGAAGTGGTGAAGA			
QY	181	CCAGTTTCATCTCATGAGGGTACATGATGGATTTGACAGCCCATCAGGCGGGTGG			
Db	1140	CCAGTTTCATCTCATGAGGGTACATGATGGATTTGACAGCCCATCAGGCGGGTGG			
QY	241	TAATATTCTACAAATGGGAAGTAAAGTGAAGTGGACAGGAAAATATCAGAAGATGAAGA			
Db	1200	TAATATTCTACAAATGGGAAGTAAAGTGAAGTGGACAGGAAAATATCAGAAGATGAAGA			
QY	301	AACGTAAGTACAAGAGCAGATGAATCTCCTAAATCAAGATGGGAATGCCTCAGGCTAGC			
Db	1260	AACGTAAGTACAAGAGCAGATGAATCTCCTAAATCAAGATGGGAATGCCTCAGGCTAGC			
QY	361	TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAACT			
Db	1320	TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAACT			
QY	421	GAAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGAAACAGGAAAATGGAGGAAGA			
Db	1380	GAAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGAAACAGGAAAATGGAGGAAGA			
QY	481	GCCTCTTGACCTGATCTTGAAGACCTAAACGCCAAGTACAACACATTAAGGTGCTTCA			
Db	1440	GCCTCTTGACCTGATCTTGAAGACCTAAACGCCAAGTACAACACATTAAGGTGCTTCA			
QY	541	AGAAGATCTAGAACAAAGCAAGTCAAGGTCAATCTCTCACTCACATGGTGGTAGT			
Db	1500	AGAAGATCTAGAACAAAGCAAGTCAAGGTCAATCTCTCACTCACATGGTGGTAGT			
QY	601	TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATGGG			
Db	1560	TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATGGG			
QY	661	AGATCGATGGGCAACATCTGTAGATGGACAGAGACCGCTGGGTCTTTACAAGACAT			
Db	1620	AGATCGATGGGCAACATCTGTAGATGGACAGAGACCGCTGGGTCTTTACAAGACAT			
QY	721	CTTCTCAATGGCAACGCTTACTGAAGAACAGTGCCTTTTAGTGCAATGGCTTCTTCA			
Db	1680	GCCTGAC-----CTAGCTCTGGACTGACCACTATTGGAGCCCTCTCTTCACTCA			
QY	781	AAAGAAAGATGCAGTGAACAAAGATTCACACAACTGGCTTTAAAGATCAAAATGAATGTT			
Db	1728	GACTGTTACTCTGGTGACACA-----ACCTGTGGTTACTAAGGAAACTGCCATCT-1777			
QY	841	ATCAAGTCTTCAAAACTGGCCGTTTAAAGCGGATCTAGAAAAGAAAAGCAATCCAT			
Db	1778	-----			
QY	901	GGGCAAACTGATTTCACTCAAAACAAAGATCTTCTTCAACACTGAAGAAATAAGTCAGTGAC			
Db	1778	-----			
QY	961	CCAGAGACGGAAGCATGGCTGGATACTTTGCCCGGTGTTGGGATAATTTAGTCCAAA			
Db	1813	-----			
QY	1021	ACTTGAAAAGAGTACAGCACAGACTCATAGATTACTGCAACAGTTCCCCCTGGACCTGGA			
Db	1021	ACTTGAAAAGAGTACAGCACAGACTCATAGATTACTGCAACAGTTCCCCCTGGACCTGGA			

us - 09 - 845 - 416 - 2_copy_960_3000.rng

Mon Apr 28 09:28:32 2003

AAD37256		AAD37256 standard; DNA; 4966 BP.	
ID	XX	AAD37256;	
AC	XX	21-AUG-2002 (first entry)	
DT	XX	Adeno-associated virus vector plasmid, AAV-MCK-delta3990.	
DE	XX	Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;	
XX	XX	adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;	
KW	XX	Becker muscular dystrophy; ds.	
KW	XX	Chimeric - Homo sapiens.	
XX	OS	Chimeric - Unidentified.	
OS	XX	WO200183695-A2.	
PN	XX	08-NOV-2001.	
XX	PD	27-APR-2001; 2001WO-US13677.	
XX	PF	28-APR-2000; 2000US-200777P.	
XX	PR	(XIAO/) XIAO X.	
XX	PA	Xiao X;	
XX	PI	WPI; 2002-049342/06.	
XX	XX	New dystrophin minigene for treating Duchenne or Becker muscular	
PS	XX	dystrophy comprises an N-terminal domain or modified N-terminal domain,	
XX	CC	rod repeats, H1 and H4 domains and a cysteine rich domain of a	
CC	CC	dystrophin gene -	
CC	CC	Example 1; Page 59-60; 71pp; English.	
CC	CC	The present invention relates to an isolated nucleotide sequence encoding	
CC	CC	a dystrophin minigene. The minigene comprises N-terminal or modified	
CC	CC	N-terminal domains, rod repeats, utrophin or a spectrin gene. The	
CC	CC	domains and cysteine-rich domains of dystrophin or utrophin genes. The	
CC	CC	invention also relates to a recombinant adeno-associated virus (AAV)	
CC	CC	comprising dystrophin minigene operably linked to an expression control	
CC	CC	element. The dystrophin minigene in operable linkage with an expression	
CC	CC	control element, in a recombinant adeno-associated virus or retrovirus is	
CC	CC	useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular	
CC	CC	dystrophy (BMD) in a mammalian subject. The present sequence is AAV	
CC	CC	vector plasmid construct containing human dystrophin minigenes, a muscle	
CC	CC	creatine kinase (MCK) promoter and a small polyA signal sequence.	
XX	XX	Sequence 4966 BP; 1403 A; 1216 C; 1232 G; 1115 T; 0 other;	
SQ	XX	Query Match 74.5%; Score 1519.8; DB 24; Length 4966;	
	XX	Best Local Similarity 87.5%; Pred. No. 0;	
	XX	Matches 1786; Conservative 0; Mismatches 72; Indels 183; Gaps 4;	
QY	1	TCCTTCACAGCATTTGGAAGCTCCTGAAGACAGCTTTAGAACAGTATTCGTTGCTTCTTC 120	
Db	1717	TCCTTCACAGCATTTGGAAGCTCCTGAAGACAGCTTTAGAACAGTATTCGTTGCTTCTTC 1836	
QY	61	TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAACAGTATTCGTTGCTTCTTC 180	
Db	1777	TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAACAGTATTCGTTGCTTCTTC 1896	
QY	121	TGCTGAGGACACATTCGAAGCAGCAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 240	
Db	1837	TGCTGAGGACACATTCGAAGCAGCAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 1956	
QY	181	CCAGTTTCATACATCAGGGGTACATGATGATTTGACAGCCCCATCAGGGCGGTTGG 300	
Db	1897	CCAGTTTCATACATCAGGGGTACATGATGATTTGACAGCCCCATCAGGGCGGTTGG 300	
QY	241	TAATATTCTACAATTTGGGAAGTAAGCTGATTGGACAGGAAATTTATCAGAAGATGAAGA	

Db 1957 TAATAATCTTACAAATGGGAAGTAAGCTGATTGGAACAGGAAATATATCAGAAGATGAAGA 2016
QY 301 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGATGGGAATGCCTCAGGGTAGC 360
Db 2017 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGATGGGAATGCCTCAGGGTAGC 2076
QY 361 TAGCATGGAAAAACAAGCAATTTACATAGAGTTTAAATGGATCTCCAGATCAGAAACT 420
Db 2077 TAGCATGGAAAAACAAGCAATTTACATAGAGTTTAAATGGATCTCCAGATCAGAAACT 2136
QY 421 GAAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 480
Db 2137 GAAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2196
QY 481 GCCTCTGGACCTGATCTTGAAGACCTTAAACGCCAAGTACAAACAACATAAAGTGCTTCA 540
Db 2197 GCCTCTGGACCTGATCTTGAAGACCTTAAACGCCAAGTACAAACAACATAAAGTGCTTCA 2256
QY 541 AGAAGATCTAGAACAAAGCAAGTCAAGGTCAATTTCTCTCACTCACATGGTGGTGTAGT 600
Db 2257 AGAAGATCTAGAACAAAGCAAGTCAAGGTCAATTTCTCTCACTCACATGGTGGTGTAGT 600
QY 601 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAGAAACAACCTTAAGGTATGGG 660
Db 2317 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAGAAACAACCTTAAGGTATGGG 2376
QY 661 AGATCGATGGGCAACATCTGTAGATGGACAGAACCGCTGGGTCTTTTACAAGACAT 720
Db 2377 AGATCGATGGGCAACATCTGTAGATGGACAGAACCGCTGGGTCTTTTACAAGACAT 720
QY 721 CCTCTCAAAATGGCAACGCTTACTGAAGAACAGTGCCTTTTGTAGTGCATGGCTTTCAGA 780
Db 2437 GCCTGAC-----CTAGCTCCTGGACTGACCACTATTTGAGCGCTCTCCTACTCA 2484
QY 781 AAAAGAAGATGCAGTGAACAAGATTCACACAACCTGCTTTTAAAGATCAAAATGAATGTT 840
Db 2485 GACTGTTACTCTGCTGACACA-----ACCTGGTGTACTAAGGAAACTGCCATCT- 2534
QY 841 ATCAAGTCTTCAAAAACCTGGCCGTTTAAAGCGGATCTAGAAAAGAAAGCAATCCAT 900
Db 2535 -----
QY 901 GGGCAACTGTATTCACTCAAACAAGATCTTCTTCAACACTGAAGAATAAGTCAGTGAC 960
Db 2535 -----CCAACTAGAAATGCCATCTTCTTGTAGTTGGAG----- 2569
QY 961 CCAGAAGACGGAAGCATGGCTGGATAACTTTGCCCGGTGTGGGATAATTTAGTCCAAA 1020
Db 2570 -----
QY 1021 ACTTGAAAAGAGTACAGCACAGACTCATAGATTACTGCAACAGTTTCCCTCGACCTGGA 1080
Db 2570 -----GTACCTACTCATAGATTACTGCAACAGTTTCCCTCGACCTGGA 2613
QY 1081 AAAGTTTCTTGCCTTACAGAAGCTGAACAACACTGCCAATGTCTTACAGGATGCTAC 1140
Db 2614 AAAGTTTCTTGCCTTACAGAAGCTGAACAACACTGCCAATGTCTTACAGGATGCTAC 2673
QY 1141 CCGTAAGGAAGGCTCCTAGAAAGCTCCAAGGGAGTAAAGAGCTGATGAACAATGGCA 1200
Db 2674 CCGTAAGGAAGGCTCCTAGAAAGCTCCAAGGGAGTAAAGAGCTGATGAACAATGGCA 2733
QY 1201 AGACCTCCAAGGTGAATGAAGCTCACACAGATGTTTATCACAACCTGGATGAACAACAG 1260
Db 2734 AGACCTCCAAGGTGAATGAAGCTCACACAGATGTTTATCACAACCTGGATGAACAACAG 2793
QY 1261 CCAAAAAATCCTGAGATCCCTGGAAGGTTCCGATGATGCAGTCTGTTACAAAAGACGTTT 1320
Db 2794 CCAAAAAATCCTGAGATCCCTGGAAGGTTCCGATGATGCAGTCTGTTACAAAAGACGTTT 2853
QY 1321 GGATAACATGAACCTCAAGTGGAGTGAACCTTCGGAAGAAAGTCTCTCAACATTAGGTCCCA 1380
Db 1380 GGATAACATGAACCTCAAGTGGAGTGAACCTTCGGAAGAAAGTCTCTCAACATTAGGTCCCA 1380

Db 2854 GGATAACATGAACCTCAAGTGGAGTGAACCTTCGGAAGAAAGTCTCTCAACATTAGGTCCCA 2913
QY 1381 TTTGGAAGCCAGTTCTGACCAAGTGAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGT 1440
Db 2914 TTTGGAAGCCAGTTCTGACCAAGTGAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGT 2973
QY 1441 GTGGCTACAGCTGAAAGATGATGAATTAAGCCGCGCAGGCACCTATTGGAGGCGACTTTC 1500
Db 2974 GTGGCTACAGCTGAAAGATGATGAATTAAGCCGCGCAGGCACCTATTGGAGGCGACTTTC 3033
QY 1501 AGCAGTTTCAGAGCAGAACGATGTACATAGGGCTTCAAGAGGGAATTTGAAACTAAAGA 1560
Db 3034 AGCAGTTTCAGAGCAGAACGATGTACATAGGGCTTCAAGAGGGAATTTGAAACTAAAGA 3093
QY 1561 ACCTGTATCATGAGTACTCTTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGA 1620
Db 3094 ACCTGTATCATGAGTACTCTTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGA 3153
QY 1621 AGSACTAGAGAAACTCTACCAGAGCCCGCAGAGAGTGCCTCCTGAGGAGAGAGCCAGAA 1680
Db 3154 AGSACTAGAGAAACTCTACCAGAGCCCGCAGAGAGTGCCTCCTGAGGAGAGAGCCAGAA 3213
QY 1681 TGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAATTTGAA 1740
Db 3214 TGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAATTTGAA 3273
QY 1741 CCTGCACCTCCGCTGACTGGCAGAGAAATAAGATGAGACCTTGAAGAGACTCCAGGAATC 1800
Db 3274 CCTGCACCTCCGCTGACTGGCAGAGAAATAAGATGAGACCTTGAAGAGACTCCAGGAATC 1860
QY 1801 TCAAGAGCCCGCAGGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGTGAAGAGTCA 1920
Db 3334 TCAAGAGCCCGCAGGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGTGAAGAGTCA 3393
QY 1861 CTGGCAGCCCGTGGGCGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGAGAAAGTCAA 1920
Db 3394 CTGGCAGCCCGTGGGCGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGAGAAAGTCAA 3453
QY 1921 GGCACCTTCGAGGAGAAATTTGGCCTCTGAAAGAGAGACGTGAGCCACGTCAATGACCTTC 1980
Db 3454 GGCACCTTCGAGGAGAAATTTGGCCTCTGAAAGAGAGACGTGAGCCACGTCAATGACCTTC 3513
QY 1981 TCGCCAGCTTACCACCTTTGGGCTTTCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGA 2040
Db 3514 TCGCCAGCTTACCACCTTTGGGCTTTCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGA 3573
QY 2041 C 2041
Db 3574 C 3574

RESULT 4

AAD37262

ID AAD37262 standard; DNA; 4990 BP.

AC AAD37262;

XX

DT 21-AUG-2002 (first entry)

XX

DE Adeno-associated virus (AAV) vector plasmid, AAV-CMV-delta3990.

XX Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.

XX Chimeric - Homo sapiens.

OS Chimeric - Cytomegalovirus.

XX Chimeric - Unidentified.

PN WO200183695-A2.

XX

PD 08-NOV-2001.

XX

Mon Apr 28 09:28:32 2003

*

27-APR-2001: 2001WO-US13677.
28-APR-2000: 2000US-200777P.
(XIAO/) XIAO X.
Xiao X;
WPI: 2002-049342/06.
New dystrophin minigene for treating Duchenne or Becker muscular dystrophy comprises an N-terminal domain or modified N-terminal domain, rod repeats, H1 and H4 domains and a cysteine rich domain of a dystrophin gene -
Example 1; Page 67-68; 71pp; English.
The present invention relates to an isolated nucleotide sequence encoding a dystrophin minigene. The minigene comprises N-terminal or modified N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control element. The dystrophin minigene in operable linkage with an expression control element, in a recombinant adeno-associated virus or retrovirus is useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is AAV vector plasmid construct containing human dystrophin minigenes, a cytomegalovirus (CMV) promoter and a small polyA signal sequence.
Sequence 4990 BP; 1439 A; 1185 C; 1208 G; 1158 T; 0 other;
Query Match 74.5%; Score 1519.8; DB 24; Length 4990;
Best Local Similarity 87.5%; Pred. NO. 0;
Matches 1786; Conservative 0; Mismatches 72; Indels 183; Gaps 4;
1 TCCTTCACAGCATTTGGAAGCTCTGAAGACAAAGTCAATTTGGCAGTTTCATTGATGGAGAG 60
1741 TCCTTCACAGCATTTGGAAGCTCTGAAGACAAAGTCAATTTGGCAGTTTCATTGATGGAGAG 1800
61 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGATATATATCGTGGCTTCTTC 120
1801 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGATATATATCGTGGCTTCTTC 1860
121 TGCTGAGGACACATTTGGAAGCTCTGAAGACAAAGTCAATTTGGCAGTTTCATTGATGGAGAG 180
1861 TGCTGAGGACACATTTGGAAGCTCTGAAGACAAAGTCAATTTGGCAGTTTCATTGATGGAGAG 1920
181 CCAGTTTCATCTACTCATGAGGGGTACATGATGGATTGACAGCCCATCAGGGCCGGTTGG 240
1921 CCAGTTTCATCTACTCATGAGGGGTACATGATGGATTGACAGCCCATCAGGGCCGGTTGG 1980
241 TAATATTTACAATTTGGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 300
1981 TAATATTTACAATTTGGAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAGTAAAG 2040
301 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTTCAAGATGGGAATGCCCTCAGGGTAGC 360
2041 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTTCAAGATGGGAATGCCCTCAGGGTAGC 2100
361 TAGCATGGAAAAACAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT 420
2101 TAGCATGGAAAAACAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT 2160
421 GAAAGAGTTGAATGACTGGCTTAACAAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 480
2161 GAAAGAGTTGAATGACTGGCTTAACAAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2220
481 GCCTCTTGGACCTGATCTTGAACACCTAAACGCCCAAGTACAACAACATGAAGTGCCTTCA 540
2221 GCCTCTTGGACCTGATCTTGAAGACCTAAACGCCCAAGTACAACAACATGAAGTGCCTTCA 2280
541 AGAAGATCTAGAAACAAGAACAAAGTCAGGCTCAATTTCTCTCACTCATCGTGGTGGTAGT 600

Db 2281 AGAAGATCTAGAAACAAGAACAAAGTCAAGGTCATCTCTCACTCATCGTGGTGGTAGT 2340
QY 601 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGG 660
Db 2341 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGG 2400
QY 661 AGATCGATGGGCAAAACATCTGTAGATGGACAGAACCGCTGGGTTCTTTTACAAGACAT 720
Db 2401 AGATCGATGGGCAAAACATCTGTAGATGGACAGAACCGCTGGGTTCTTTTACAAGACCA 2460
QY 721 CCTTCTCAAAATGGCAACGCTTACTGAAGAACAGTGCCTTTTCTAGTCATGGCTTTTCCA 780
Db 2461 GCCTGAC-----CTAGCTCTGGACTGACCACTATTGGAGCCTCTCTCTACTCA 2508
QY 781 AAAAGAAGATGCAGTGAACAAGATTCACACAACCTGGCTTTTAAAGATCAAAATGAATGTT 840
Db 2509 GACTGTTACTCTGGTGACACA-----ACCTGTGGTTACTAAGGAACCTGCCATCT- 2558
QY 841 ATCAAGTCTTCAAAACAGTGGCCGTTTAAAGAGCGGATCTAGAAAAGAAAAGCAATCCAT 900
Db 2559 ----- 2558
QY 901 GGGCAAACTGTATTCACTCAACAAGATCTTCTTTCAACACTGAAGAATAAGTCAGTGAC 960
Db 2559 ----- 2593
QY 961 CCAGAAGACGGAAGCATGGCTGGATAAATTTGCCGGTGTGGGATAATTTAGTCCAAA 1020
Db 2594 ----- 2593
QY 1021 ACTTGAAAAGAGTACAGACAGACTCATAGATTAAGTCAACAGTTTCCCTCCCTGGACCTGGA 1080
Db 2594 -----GTACCTACTCATAGATTACTGCAACAGTTTCCCTCCCTGGACCTGGA 2637
QY 2638 AAAGTTTCTTGGCTGGCTTACAGAAGCTGAAACAACTGCCAATGTCTTACAGGATGCTAC 2697
Db 1141 CCCTAAGGAAAGGCTCCTTAGAAGACTCCAAAGGAGTAAAGAGAGTGAATGAACAAATGGCA 1200
QY 2698 CCGTAAGGAAAGGCTCCTTAGAAGACTCCAAAGGAGTAAAGAGAGTGAATGAACAAATGGCA 1260
Db 1201 AGACCTCCAAAGGTGAAATTTGAAGCTCACACAGATGTTTATCACAACCTGGATGAAACACAG 1260
QY 2758 AGACCTCCAAAGGTGAAATTTGAAGCTCACACAGATGTTTATCACAACCTGGATGAAACACAG 2817
Db 1261 CCAAAAAATCCTGAGATCCCTGGAAAGTTCCGATGATGAGTCCCTGTTTACAAGACGTTT 2877
QY 2818 CCAAAAAATCCTGAGATCCCTGGAAAGTTCCGATGATGAGTCCCTGTTTACAAGACGTTT 2937
Db 1321 GGATAACATGAACCTTCAAGTGGAGTGAACCTTCGGAAGAAAGTCTCTCAACATTTAGTCCCA 1380
QY 2878 GGATAACATGAACCTTCAAGTGGAGTGAACCTTCGGAAGAAAGTCTCTCAACATTTAGTCCCA 1440
Db 1381 TTTGGAAAGCCAGTTCTGACCCAGTGAAGCGTCTGCACCTTTCTCTGACGGAACCTTCTGTT 1440
QY 2938 TTTGGAAAGCCAGTTCTGACCCAGTGAAGCGTCTGCACCTTTCTCTGACGGAACCTTCTGTT 2997
Db 1441 GTGGCTACAGCTGAAGAAGATGATGAATTAAGCCGCGCAGGCACCTATTGGAGCGGACTTTCC 1500
QY 2998 GTGGCTACAGCTGAAGAAGATGATGAATTAAGCCGCGCAGGCACCTATTGGAGCGGACTTTCC 3057
Db 1501 AGCAGTTTCAAGAGCAGAACGATGTACATAGGGCCCTTCAAGAGGGAATTTGAAACCTAAAGA 1560
QY 3058 AGCAGTTTCAAGAGCAGAACGATGTACATAGGGCCCTTCAAGAGGGAATTTGAAACCTAAAGA 3117
Db 1561 ACCTGTAATCATGAGTACTCTTACAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGA 1620
QY 3118 ACCTGTAATCATGAGTACTCTTACAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGA 3177
Db 1621 AGGACTAGAGAAACTCTACCAGGAGCCGACAGAGCTGCCTCCTGAGGAGAGAGCCAGAA 1680
QY 1621 AGGACTAGAGAAACTCTACCAGGAGCCGACAGAGCTGCCTCCTGAGGAGAGAGCCAGAA 1680

Db 3178 AGGACTAGAGAAACTCTACAGGAGCCAGAGAGCTGCCTCTGAGGAGAGAGCCAGAA 3237
QY 1681 TGTCACCTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTGAA 1740
Db 3238 TGTCACCTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTGAA 1740
QY 1741 CCTGCACCTCCGCTGACTGGCAGAGAGAAAAATAGATGAGACCCCTTGAAGACTCCAGGAAC 3297
Db 3298 CCTGCACCTCCGCTGACTGGCAGAGAGAAAAATAGATGAGACCCCTTGAAGACTCCAGGAAC 1800
QY 1801 TCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGATC 3357
Db 3358 TCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGATC 1860
QY 1861 CTGGCAGCCCGTGGCGGATCTCTCTATTGACTCTCTCCCAAGATCACCTCGAGAAAGTCAA 3417
Db 3418 CTGGCAGCCCGTGGCGGATCTCTCTATTGACTCTCTCCCAAGATCACCTCGAGAAAGTCAA 1920
QY 1921 GGCACCTTCGAGGAGAAATTGGCCCTCTGAAAGAGAACGCTGAGCCACGTCATGACCTTGC 3477
Db 3478 GGCACCTTCGAGGAGAAATTGGCCCTCTGAAAGAGAACGCTGAGCCACGTCATGACCTTGC 1980
QY 1981 TCGCCAGCTTACCACTTTGGGCATTGAGCTCTCACCGTATACCTCAGCACTCTGGAAGA 3537
Db 3538 TCGCCAGCTTACCACTTTGGGCATTGAGCTCTCACCGTATACCTCAGCACTCTGGAAGA 2040
QY 2041 C 2041
Db 3598 C 3598

RESULT 5
AAD37237
ID AAD37237 standard; DNA; 3858 BP.
XX
AC AAD37237;
XX
DT 21-AUG-2002 (first entry)
XX
DE Human dystrophin minigene delta3849.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Homo sapiens.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
XX
PI Xiao X;
XX
WPI; 2002-049342/06.
XX
PT New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -

Example 1; Page 48-49; 71pp; English.
The present invention relates to an isolated nucleotide sequence encoding
a dystrophin minigene. The minigene comprises N-terminal or modified
N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
domains and cysteine-rich domains of dystrophin or utrophin genes. The
invention also relates to a recombinant adeno-associated virus (AAV)

comprising dystrophin minigene operably linked to an expression control
element. The dystrophin minigene in operable linkage with an expression control
control element, in a recombinant adeno-associated virus or retrovirus is
useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
dystrophy (BMD) in a mammalian subject. The present sequence is human
dystrophin minigene delta3849 containing nucleotides 1-1668 (N-terminus,
hinge H1 and rods R1, R2), 8059-10227 (rods R22, R23 and R24, hinge H4
and CR domain) and 11047-11058 (dystrophin last 3 amino acids).

SQ Sequence 3858 BP; 1189 A; 866 C; 905 G; 898 T; 0 other;
Query Match 67.8%; Score 1383; DB 24; Length 3858;
Best Local Similarity 84.1%; Pred. No. 0;
Matches 1717; Conservative 0; Mismatches 0; Indels 324; Gaps 1;

QY 1 TCCTTCACAGCATTGGGAGCTCTGAAGACAAGTCAATTTGGCAGTTCATTGATGGAGAG 60
Db 960 TCCTTCACAGCATTGGGAGCTCTGAAGACAAGTCAATTTGGCAGTTCATTGATGGAGAG 60
QY 61 TGAAGTAAACCTGGACCGTTATCAACACAGCTTTAGAGAAAGTATTATCGTGGCTTCTTC 1019
Db 1020 TGAAGTAAACCTGGACCGTTATCAACACAGCTTTAGAGAAAGTATTATCGTGGCTTCTTC 120
QY 121 TGCTGAGGACACATTCGAAGCACAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 1079
Db 1080 TGCTGAGGACACATTCGAAGCACAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 180
QY 181 CCAGTTTCATCTCATGAGGGGTACATGATGATTTGACAGCCCATCAGGCGCGGTGG 1139
Db 1140 CCAGTTTCATCTCATGAGGGGTACATGATGATTTGACAGCCCATCAGGCGCGGTGG 240
QY 241 TAATATTCTACAATTGGGAAGTAAGCTGATTGGAACAGGAAATATATCAGAAGATGAAGA 1199
Db 1200 TAATATTCTACAATTGGGAAGTAAGCTGATTGGAACAGGAAATATATCAGAAGATGAAGA 300
QY 301 AACTGAAGTACAAGAGCAGATGAATCTCTAAATCAAGATGGGAATGCCTCAGGGTAGC 1259
Db 1260 AACTGAAGTACAAGAGCAGATGAATCTCTAAATCAAGATGGGAATGCCTCAGGGTAGC 360
QY 361 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGATCTCCAGATCAGAAAACT 1319
Db 1320 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGATCTCCAGATCAGAAAACT 420
QY 421 GAAAGAGTTGAATGACTGGCTAACCAAAACAGAGAAAGAAAGAAAGAAAGAAAGAAAG 1379
Db 1380 GAAAGAGTTGAATGACTGGCTAACCAAAACAGAGAAAGAAAGAAAGAAAGAAAGAAAG 480
QY 481 GCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCAAGTACAAACAAACATAAGGTCTTCA 1439
Db 1440 GCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCAAGTACAAACAAACATAAGGTCTTCA 540
QY 541 AGAAGATCTAGAACAAAGCAAGTCAAGGTCAATCTCTCACTCACATGGTGGTAGT 1499
Db 1500 AGAAGATCTAGAACAAAGCAAGTCAAGGTCAATCTCTCACTCACATGGTGGTAGT 600
QY 601 TGATGAATCTAGTGAGATCACGCAACTGCTGCTTTGGAAGAAACAACTTAAGGTATTGGG 1559
Db 1560 TGATGAATCTAGTGAGATCACGCAACTGCTGCTTTGGAAGAAACAACTTAAGGTATTGGG 660
QY 661 AGATCGATGGGCAACATCTGTAGATGGACAGAACGCCGTGGGTCTTTTACAAAGACAT 1619
Db 1620 AGATCGATGGGCAACATCTGTAGATGGACAGAACGCCGTGGGTCTTTTACAAAGACAT 720
QY 721 CCTTCTCAAAATGGCAACGCTTACTGAAGAACAGTGCCTTTTAGTGCATGGCTTTCAGA 1677
Db 1678 -----
QY 781 AAAAGAAGATGCAGTGAACAGATTCACACAACCTGCTTTAAAGATCAAAATGAATGTT 1677
Db 1678 -----
QY 841 ATCAAGTCTTCAAAACTGGCCGTTTTTAAAGCGGATCTAGAAAAGAAAAAGCAATCCAT 900

Mon Apr 28 09:28:32 2003

Db	1678	-----	1677
QY	901	GGGCAAACTGTATTCACTCAAAACAAGATCTTCTTTCAACACTGAAGAATAAGTCAGTGAC	960
Db	1678	-----	1677
QY	961	CCAGAACGGAAGCATGGCTGGATAACTTTGCCCGGTGTTGGGATAATTTAGTCCAAAA	1020
Db	1678	-----	1677
QY	1021	ACTTGAAGAGAGTACAGCACAGACTCATAGATTACTGCAACAGTTTCCCTGGACCTGGA	1080
Db	1678	-----	1677
QY	1081	AAAGTTTCTTGCTGGCTTACAGAAGCTGAACAACTGCCAATGTCTACAGGATGCTAC	1140
Db	1716	AAAGTTTCTTGCTGGCTTACAGAAGCTGAACAACTGCCAATGTCTACAGGATGCTAC	1775
QY	1141	CCGTAAGGAAAGGCTCCTAGAAGACTCCAAGGGAGTAAAGAGCTGATGAAACAATGGCA	1200
Db	1776	CCGTAAGGAAAGGCTCCTAGAAGACTCCAAGGGAGTAAAGAGCTGATGAAACAATGGCA	1835
QY	1201	AGACCTCCAAGGTTGAAATTTGAAGCTCACACAGATGTTTATCAACAACTGGATGAAACAG	1260
Db	1836	AGACCTCCAAGGTTGAAATTTGAAGCTCACACAGATGTTTATCAACAACTGGATGAAACAG	1895
QY	1261	CAAAAAATCCTGAGATCCCTGGAGGTTCCGATGATGCAGTCTCTCAACATTAGTCCCA	1320
Db	1896	CAAAAAATCCTGAGATCCCTGGAGGTTCCGATGATGCAGTCTCTCAACATTAGTCCCA	1955
QY	1321	GGATAACATGAACCTCAAGTGGAGTGAACCTCGGAAAAAGTCTCTCAACATTAGTCCCA	1380
Db	1956	GGATAACATGAACCTCAAGTGGAGTGAACCTCGGAAAAAGTCTCTCAACATTAGTCCCA	2015
QY	1381	TTTGGAGCCAGTTCTGACCAAGTGAAGCGTCTGCACCTTCTCTGCAGGAACTTCTGCT	1440
Db	2016	TTTGGAGCCAGTTCTGACCAAGTGAAGCGTCTGCACCTTCTCTGCAGGAACTTCTGCT	2075
QY	1441	GTGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTTGGAGCGACTTTC	1500
Db	2076	GTGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTTGGAGCGACTTTC	2135
QY	1501	AGCAGTTTCAGAACGACGATGTACATAGGCGCTTCAAGAGGGATTTGAAAACTAAAGA	1560
Db	2136	AGCAGTTTCAGAACGACGATGTACATAGGCGCTTCAAGAGGGATTTGAAAACTAAAGA	2195
QY	1561	ACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGA	1620
Db	2196	ACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGA	2255
QY	1621	AGGACTAGAGAAACTCTACAGAGAGCCAGAGAGCTGCCTCTGAGGAGAGAGCCAGAA	1680
Db	2256	AGGACTAGAGAAACTCTACAGAGAGCCAGAGAGCTGCCTCTGAGGAGAGAGCCAGAA	2315
QY	1681	TGTCACCTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTGA	1740
Db	2316	TGTCACCTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTGA	2375
QY	1741	CCTGCACCTCCGCTGACTGGCAGAGAAAATAGATGAGACCCCTTGAAGAACTCCAGGA	1800
Db	2376	CCTGCACCTCCGCTGACTGGCAGAGAAAATAGATGAGACCCCTTGAAGAACTCCAGGA	2435
QY	1801	TCAAGAGGCCACGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATC	1860
Db	2436	TCAAGAGGCCACGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATC	2495
QY	1861	CTGGCAGCCCGTGGCGATCTCCTCATTTGACTCTCTCAAGATCACCTCGAGAAAGTCAA	1920
Db	2496	CTGGCAGCCCGTGGCGATCTCCTCATTTGACTCTCTCAAGATCACCTCGAGAAAGTCAA	2555
QY	1921	GGCACTTCGAGGAGAAATTGCGCTCTGAAAGAGAACGTCAGCCACGTCAATGACCTTGC	1980
Db	2556	GGCACTTCGAGGAGAAATTGCGCTCTGAAAGAGAACGTCAGCCACGTCAATGACCTTGC	2615
QY	1981	TCGCCAGCTTACCACCTTTGGCATTCAGCTCTCACCCTGATATACCTCAGCACTCTGGA	2040
Db	2616	TCGCCAGCTTACCACCTTTGGCATTCAGCTCTCACCCTGATATACCTCAGCACTCTGGA	2675
QY	2041	C 2041	
Db	2676	C 2676	
RESULT 6			
AAD37257			
ID	AAD37257 standard; DNA; 4825 BP.		
XX	AAD37257;		
AC	AAD37257;		
XX	21-AUG-2002 (first entry)		
DT	Adeno-associated virus vector plasmid, AAV-MCK-delta3849.		
XX	Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;		
KW	adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;		
KW	Becker muscular dystrophy; ds.		
XX	Chimeric - Homo sapiens.		
OS	Chimeric - Unidentified.		
XX	WO200183695-A2.		
PN	08-NOV-2001.		
XX	27-APR-2001; 2001WO-US13677.		
PF	28-APR-2000; 2000US-200777P.		
XX	(XIAO/) XIAO X.		
PA	Xiao X;		
PI	WPI; 2002-049342/06.		
XX	New dystrophin minigene for treating Duchenne or Becker muscular		
DR	dystrophy comprises an N-terminal domain or modified N-terminal domain,		
XX	rod repeats, H1 and H4 domains and a cysteine rich domain of a		
PT	dystrophin gene		
PT	Example 1; Page 61-62; 71pp; English.		
XX	The present invention relates to an isolated nucleotide sequence encoding		
CC	a dystrophin minigene. The minigene comprises N-terminal or modified		
CC	N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4		
CC	domains and cysteine-rich domains of dystrophin or utrophin genes. The		
CC	invention also relates to a recombinant adeno-associated virus (AAV)		
CC	comprising dystrophin minigene operably linked to an expression control		
CC	element. The dystrophin minigene in operable linkage with an expression		
CC	control element, in a recombinant adeno-associated virus or retrovirus is		
CC	useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular		
CC	dystrophy (BMD) in a mammalian subject. The present sequence is AAV		
CC	vector plasmid construct containing human dystrophin minigenes, a muscle		
CC	creatine kinase (MCK) promoter and a small polyA signal sequence.		
XX	Sequence 4825 BP; 1369 A; 1175 C; 1204 G; 1077 T; 0 other;		
SQ	Query Match 67.8%; Score 1383; DB 24; Length 4825;		
	Best Local Similarity 84.1%; Pred. No. 0;		
	Matches 1717; Conservative 0; Mismatches 0; Indels 324; Gaps 1;		
QY	1	TCCTTCACAGCATTTGGAGCTCCTGAAGACAAGTCATTTGGCAGTTTCATTGATGGAGAG	60
Db	1717	TCCTTCACAGCATTTGGAGCTCCTGAAGACAAGTCATTTGGCAGTTTCATTGATGGAGAG	1776
QY	61	TGAAGTAAACCTGGACCGTTATCAACAGCTTTAGAGAAAGTATTATCGTGGCTTCTTC	120

Db 1777 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAGAAGTATATATCGTGGCTTCTTTC 1836
QY 121 TGCTGAGGACACATTGCAAGCACACAGGAGAGATTCTTAATGATGTGGAAGTGGTGAAGA 180
Db 1837 TGCTGAGGACACATTGCAAGCACACAGGAGAGATTCTTAATGATGTGGAAGTGGTGAAGA 1896
QY 181 CCAGTTTCACTACTCATGAGGGTACATGATGGATTGACAGCCCCATCAGGGCCGGGTGG 240
Db 1897 CCAGTTTCACTACTCATGAGGGTACATGATGGATTGACAGCCCCATCAGGGCCGGGTGG 1956
QY 241 TAATATTCTACAATTGGGAAGTAAGCTGATTGGACAGGAAAATTATCAGAAGATGAAGA 300
Db 1957 TAATATTCTACAATTGGGAAGTAAGCTGATTGGACAGGAAAATTATCAGAAGATGAAGA 2016
QY 301 AACTGAAGTACAGACAGATGAATCTCTAAATTCAGATGGGAATGCTCAGGATAGC 360
Db 2017 AACTGAAGTACAGACAGATGAATCTCTAAATTCAGATGGGAATGCTCAGGATAGC 2076
QY 361 TAGCATGGAAGAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAACT 420
Db 2077 TAGCATGGAAGAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAACT 2136
QY 421 GAAAGAGTTGAATGACTGGCTTAACAAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAA 480
Db 2137 GAAAGAGTTGAATGACTGGCTTAACAAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAA 2196
QY 481 GCCTCTTGACCTGATCTTGAAGACCTAAACGCCAAGTACAAACAAACATTAAGGTCTCA 540
Db 2197 GCCTCTTGACCTGATCTTGAAGACCTAAACGCCAAGTACAAACAAACATTAAGGTCTCA 2256
QY 541 AGAAGATCTAGAACAAGAACAAAGTCAAGGTCAATCTCTCACTCACATGCTGGTGTAGT 600
Db 2257 AGAAGATCTAGAACAAGAACAAAGTCAAGGTCAATCTCTCACTCACATGCTGGTGTAGT 2316
QY 601 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGG 660
Db 2317 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGG 2376
QY 661 AGATCATGGGCAAAACATCTGTAGATGGACAGAACCGCTGGGTCTTTTACAAGACAT 720
Db 2377 AGATCATGGGCAAAACATCTGTAGATGGACAGAACCGCTGGGTCTTTTACAAGACAT 2434
QY 721 CCTCTCAAATGGCAACGCTTACTGAAGAACAGTGCCTTTTAGTGCATGGCTTTCAGA 780
Db 2435 ----- 2434
QY 781 AAAAGAAGATGCAGTGAACAAGATTCAACACAACCTGGCTTTAAAGATCAAAATGAAATGTT 840
Db 2435 ----- 2434
QY 841 ATCAAGTCTTCAAAAACCTGGCCGTTTTTAAAGCGGATCTAGAAAAGAAAAGCAATCCAT 900
Db 2435 ----- 2434
QY 901 GGGCAAACTGTATTCACTCAAAACAAGATCTTCTTCAACACTGAAGAATAAGTCAAGTGAC 960
Db 2435 ----- 2434
QY 961 CCAGAAGACGGAAGCATGGCTGGATAACTTTGCCCGGTGTTGGGATAATTTAGTCCAAA 1020
Db 2435 ----- 2434
QY 1021 ACTTGAAAAGAGTACAGCAGACACTCATAGATTACTGCAACAGTTTCCCTGGACCTGGA 1080
Db 2435 -----ACTCATAGATTACTGCAACAGTTTCCCTGGACCTGGA 2472
QY 1081 AAAGTTTCTTGCCTGGCTTACAGAAGCTGAAACAACCTGCCAATGTCTTACAGGATGCTAC 1140
Db 2473 AAAGTTTCTTGCCTGGCTTACAGAAGCTGAAACAACCTGCCAATGTCTTACAGGATGCTAC 2532
QY 1141 CCGTAAGGAAGGCTCCTAGAAGACTCCAAGGGAGTAAAAGAGCTGATGAAACAATGGCA 1200
Db 2533 CCGTAAGGAAGGCTCCTAGAAGACTCCAAGGGAGTAAAAGAGCTGATGAAACAATGGCA 2592

QY 1201 AGACCTCCAAAGGTGAAATTGAAGCTCACACAGATGTTTATCACAACCTGGATGAAAACAG 1260
Db 2593 AGACCTCCAAAGGTGAAATTGAAGCTCACACAGATGTTTATCACAACCTGGATGAAAACAG 2652
QY 1261 CCAAAAAATCCTGAGATCCCTGGAAAGTTCAGATGATGCAGTCCCTGTTACAAAAGACGTTT 1320
Db 2653 CCAAAAAATCCTGAGATCCCTGGAAAGTTCAGATGATGCAGTCCCTGTTACAAAAGACGTTT 2712
QY 1321 GGATAACATGAACCTTCAAGTGGAGTGAACCTTCGAAAAAGTCTCTCAACATTAGGTCCCA 1380
Db 2713 GGATAACATGAACCTTCAAGTGGAGTGAACCTTCGAAAAAGTCTCTCAACATTAGGTCCCA 2772
QY 1381 TTTGGAAAGCCAGTCTGACCAAGTGAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGT 1440
Db 2773 TTTGGAAAGCCAGTCTGACCAAGTGAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGT 2832
QY 1441 GTGGCTACAGCTGAAAAGATGATGAATTAAGCGGCGAGGCACCTATTGGAGCGGACTTTC 1500
Db 2833 GTGGCTACAGCTGAAAAGATGATGAATTAAGCGGCGAGGCACCTATTGGAGCGGACTTTC 2892
QY 1501 AGCAGTTCAGAAAGCAGAACGATGTACATAGGGCCCTTCAAGAGGGAATTTGAAAACATAAGA 1560
Db 2893 AGCAGTTCAGAAAGCAGAACGATGTACATAGGGCCCTTCAAGAGGGAATTTGAAAACATAAGA 2952
QY 1561 ACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTTGA 1620
Db 2953 ACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTTGA 3012
QY 1621 AGGACTAGAGAAACTCTACAGGAGCCAGAGAGCTGCCTCCTGAGGAGAGAGCCAGAA 1680
Db 3013 AGGACTAGAGAAACTCTACAGGAGCCAGAGAGCTGCCTCCTGAGGAGAGAGCCAGAA 3072
QY 1681 TGTCACCTCGGCTTCTACGAAAGCAGGCTGAGGAGTCAATACTGAGTGGGAAAAATTGAA 1740
Db 3073 TGTCACCTCGGCTTCTACGAAAGCAGGCTGAGGAGTCAATACTGAGTGGGAAAAATTGAA 3132
QY 1741 CCTGCACCTCGGCTGACTGGCAGAGAAAAATAGATGAGACCTTTGAAAGACTCCAGGAAC 1800
Db 3133 CCTGCACCTCGGCTGACTGGCAGAGAAAAATAGATGAGACCTTTGAAAGACTCCAGGAAC 3192
QY 1801 TCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATC 1860
Db 3193 TCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATC 3252
QY 1861 CTGGCAGCCCGTGGCGGATCTCCTCATTGACTCTCTCCAAAGATCACCTCGAGAAAGTCAA 1920
Db 3253 CTGGCAGCCCGTGGCGGATCTCCTCATTGACTCTCTCCAAAGATCACCTCGAGAAAGTCAA 3312
QY 1921 GGCACCTCGAGGAGAAATTGCGCCTCTGAAAGAGACGTGAGCCACGTCAATGACCTTGC 1980
Db 3313 GGCACCTCGAGGAGAAATTGCGCCTCTGAAAGAGACGTGAGCCACGTCAATGACCTTGC 3372
QY 1981 TCGCCAGCTTACCACCTTTGGGCATTCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGA 2040
Db 3373 TCGCCAGCTTACCACCTTTGGGCATTCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGA 3432
QY 2041 C 2041
Db 3433 C 3433

RESULT 7
AAD37263
ID AAD37263 standard; DNA; 4848 BP.
XX
AC AAD37263;
XX

DT 21-AUG-2002 (first entry)
XX
DE Adeno-associated virus (AAV) vector plasmid, AAV-CMV-delta3849.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;

Mon Apr 28 09:28:32 2003

adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;

Becker muscular dystrophy; ds.

Chimeric - Homo sapiens.
OS Chimeric - Cytomegalovirus.
OS Chimeric - Unidentified.

XX WO200183695-A2.

PN 08-NOV-2001.

XX 27-APR-2001; 2001WO-US13677.

PF 28-APR-2000; 2000US-200777P.

XX (XIAO/) XIAO X.
XX xiao x;

XX WPI; 2002-049342/06.

DR New dystrophin minigene for treating Duchenne or Becker muscular
XX dystrophy comprises an N-terminal domain or modified N-terminal domain,
XX rod repeats, H1 and H4 domains and a cysteine rich domain of a
XX dystrophin gene -

XX Example 1; Page 68-70; 71pp; English.

PS The present invention relates to an isolated nucleotide sequence encoding
XX a dystrophin minigene. The minigene comprises N-terminal or modified
XX N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
XX domains and cysteine-rich domains of dystrophin or utrophin genes. The
XX invention also relates to a recombinant adeno-associated virus (AAV)
XX comprising dystrophin minigene operably linked to an expression control
XX element. The dystrophin minigene in operable linkage with an expression
XX control element, in a recombinant adeno-associated virus or retrovirus is
XX useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
XX dystrophy (BMD) in a mammalian subject. The present sequence is AAV
XX vector plasmid construct containing human dystrophin minigenes, a
XX cytomagalovirus (CMV) promoter and a small polyA signal sequence.

XX Sequence 4848 BP; 1405 A; 1144 C; 1180 G; 1119 T; 0 other;

XX Query Match 67.8%; Score 1383; DB 24; Length 4848;
XX Best Local Similarity 84.1%; Pred. No. 0;
XX Matches 1717; Conservative 0; Mismatches 0; Indels 324; Gaps 1;

QY 1 TCCTTCACAGCATTTGGAAGCTCCTGAAGACACAGCTTTAGAGAGATTTGGCAGTTTCATTGATGGAGAG 60
Db 1740 TCCTTCACAGCATTTGGAAGCTCCTGAAGACACAGCTTTAGAGAGATTTGGCAGTTTCATTGATGGAGAG 1799
QY 61 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAGAGATTTGGCAGTTTCATTGATGGAGAG 180
Db 1800 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAGAGATTTGGCAGTTTCATTGATGGAGAG 1859
QY 121 TGCTGAGGACACATTGCAAGCACACAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 180
Db 1860 TGCTGAGGACACATTGCAAGCACACAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 1919
QY 181 CCAGTTTCATCTACTGAGGGGTACATGATGGATTTGACAGCCCCATCAGGGCCGGTTGG 240
Db 1920 CCAGTTTCATCTACTGAGGGGTACATGATGGATTTGACAGCCCCATCAGGGCCGGTTGG 1979
QY 241 TAATATCTACAAATTTGGGAAGTAAGCTGATTGGAACAGGAAATTTATCAGAAAGATGAAGA 300
Db 1980 TAATATCTACAAATTTGGGAAGTAAGCTGATTGGAACAGGAAATTTATCAGAAAGATGAAGA 2039
QY 301 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGATGGGAATGCCTCAGGGTAGC 360
Db 2040 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGATGGGAATGCCTCAGGGTAGC 2099
QY 361 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT 420

Db 2100 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTTTAATGGATCTCCAGAAATCAGAAACT 2159
QY 421 GAAAGAGTTGAATGACTGGCTTAACAAAAACAGAAAGAAAGAAAAATGGAGGAAGA 480
Db 2160 GAAAGAGTTGAATGACTGGCTTAACAAAAACAGAAAGAAAGAAAAATGGAGGAAGA 2219
QY 481 GCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCCAAGTACAACAACATAAAGGTGCTTCA 540
Db 2220 GCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCCAAGTACAACAACATAAAGGTGCTTCA 2279
QY 541 AGAAGATCTAGAACAAAGCAAGTCAAGGTCAATCTCTCACTCACATGGTGGTGGTAGT 600
Db 2280 AGAAGATCTAGAACAAAGCAAGTCAAGGTCAATCTCTCACTCACATGGTGGTGGTAGT 2339
QY 601 TGATGAATCTAGTGGAGATCAGCAACTGCTGCTTTGGAAGAAACAACTTAAGGTATGGG 660
Db 2340 TGATGAATCTAGTGGAGATCAGCAACTGCTGCTTTGGAAGAAACAACTTAAGGTATGGG 2399
QY 661 AGATCGATGGGCAACATCTGTAGATGGACAGACCGCTGGGTCTTTTACAAGAC-- 2457
Db 2400 AGATCGATGGGCAACATCTGTAGATGGACAGACCGCTGGGTCTTTTACAAGAC-- 2457
QY 721 CCTTCTCAAAATGGCAACGCTTACTGGAAGAACAGTGCCTTTTGTGTCATGGCTTTTCAGA 780
Db 2458 -- 2457
QY 781 AAAAGAAGATGCAGTGAACAAGATTACACACAACTGGCTTTAAAGATCAAAATGAAATGTT 840
Db 2458 -- 2457
QY 841 ATCAAGTCTTCAAAACTGGCCGTTTAAAGAGCGGATCTAGAAAAAGAAAGCAATCCAT 900
Db 2458 -- 2457
QY 901 GGGCAAACTGTATTCACTCAAAACAGATCTTCTTCAACACTGAAGAATAAGTCAGTGAC 960
Db 2458 -- 2457
QY 961 CCAGAACGGAAGCATGGCTGGATAAATTTGCCGGTGTGGGATAATTTAGTCCAAAA 1020
Db 2458 -- 2457
QY 1021 ACTTGAAGAGTACAGCACAGACTCATAGATTACTGCAACAGTTCCCTCCCTGGACCTGGA 1080
Db 2458 -- 2457
QY 1081 AAAGTTTCTTGCCTGGCTTACAGAAAGCTGAAACAACTGCCAATGTCTTACAGGATGCTAC 2555
Db 2496 AAAGTTTCTTGCCTGGCTTACAGAAAGCTGAAACAACTGCCAATGTCTTACAGGATGCTAC 2555
QY 1141 CCGTAAGGAAAGGCTCCTAGAAGACTCCAAGGGAGTAAAGAGCTGATGAAACAATGGCA 1200
Db 2556 CCGTAAGGAAAGGCTCCTAGAAGACTCCAAGGGAGTAAAGAGCTGATGAAACAATGGCA 2615
QY 1201 AGACCTCAAGGTGAAATTTGAAGCTCACACAGATGTTTATCACAACTGGATGAAACAG 1260
Db 2616 AGACCTCAAGGTGAAATTTGAAGCTCACACAGATGTTTATCACAACTGGATGAAACAG 2675
QY 1261 CCAAAAAATCCTGAGATCCCTGGAAGTTCCGATGATGCAGTCCCTGTACAAAGACGTTT 1320
Db 2676 CCAAAAAATCCTGAGATCCCTGGAAGTTCCGATGATGCAGTCCCTGTACAAAGACGTTT 2735
QY 1321 GGATAACATGAACCTCAAGTGGAGTGAACCTTCGAAAAAGTCTCTCAACATAGGTCCCA 1380
Db 2736 GGATAACATGAACCTCAAGTGGAGTGAACCTTCGAAAAAGTCTCTCAACATAGGTCCCA 2795
QY 1381 TTTGGAAGCCAGTTCTGACAGTGGAGCGTCTGCACCTTCTCTGCAGGAACCTTCTGGT 1440
Db 2796 TTTGGAAGCCAGTTCTGACAGTGGAGCGTCTGCACCTTCTCTGCAGGAACCTTCTGGT 2855
QY 1441 GTGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGACCTATTTGGAGCGGACTTTC 1500
Db 2856 GTGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGACCTATTTGGAGCGGACTTTC 2915

QY 1501 AGCAGTTCAGAAAGCAGAACGATGTACATAGGGCCCTTCAAGAGGGAATGAAACTAAAGA 1560
Db 2916 AGCAGTTCAGAAAGCAGAACGATGTACATAGGGCCCTTCAAGAGGGAATGAAACTAAAGA 2975
QY 1561 ACCTGTAATCATGAGTACTCTTGTAGACTGTACGAAATTTCTGACAGAGCAGCCTTTGGA 1620
Db 2976 ACCTGTAATCATGAGTACTCTTGTAGACTGTACGAAATTTCTGACAGAGCAGCCTTTGGA 1620
QY 1621 AGGACTAGAGAACTCTTACCAGGAGCCAGAGAGCTGCCTCTCTGAGGAGAGCCAGAA 1680
Db 3036 AGGACTAGAGAACTCTTACCAGGAGCCAGAGAGCTGCCTCTCTGAGGAGAGCCAGAA 1680
QY 1681 TGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACCTGAGTGGGAAATTTGAA 1740
Db 3096 TGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACCTGAGTGGGAAATTTGAA 1740
QY 1741 CCTGCACTCGGCTGACTGGCAGAGAGAAATAGATGAGACCTTGAAGACTCCAGGAACT 1800
Db 3156 CCTGCACTCGGCTGACTGGCAGAGAGAAATAGATGAGACCTTGAAGACTCCAGGAACT 1800
QY 1801 TCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGATC 3215
Db 3216 TCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGATC 3215
QY 1861 CTGGCAGCCCTGGGCGATCTCTCATTTGACTCTCTCCAGATCACTCGAGAAAGTCAA 1920
Db 3276 CTGGCAGCCCTGGGCGATCTCTCATTTGACTCTCTCCAGATCACTCGAGAAAGTCAA 1920
QY 1921 GGCACCTCGAGGAGAAATTTGGCCTCTGAAAGAGAACGTGAGCCACGTCAATGACCTTGC 3335
Db 3336 GGCACCTCGAGGAGAAATTTGGCCTCTGAAAGAGAACGTGAGCCACGTCAATGACCTTGC 3335
QY 1981 TCGCCAGCTTACCACCTTTGGCATTGAGCTCTACCGTATACCTCAGCAGCTCTGGAAGA 2040
Db 3396 TCGCCAGCTTACCACCTTTGGCATTGAGCTCTACCGTATACCTCAGCAGCTCTGGAAGA 2040
QY 2041 C 2041
Db 3456 C 3456

RESULT 8
AAD37260
ID AAD37260 standard; DNA; 4414 BP.
XX
AC AAD37260;
XX
DT 21-AUG-2002 (first entry)
XX
DE Adeno-associated virus vector plasmid, AAV-MCK-3447.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Unidentified.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
XX 27-APR-2001; 2001WO-US13677.
XX
XX 28-APR-2000; 2000US-200777P.
XX
XX (XIAO/) XIAO X.
XX
XX Xiao X;
XX
XX WPI; 2002-049342/06.
XX

PT New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -
XX Example 1; Page 65-66; 71pp; English.
XX
CC The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is AAV
CC vector plasmid construct containing human dystrophin minigenes, a muscle
CC creatine kinase (MCK) promoter and a small polyA signal sequence.
XX Sequence 4414 BP; 1255 A; 1075 C; 1086 G; 998 T; 0 other;
SQ

Query Match
Best Local Similarity 51.6%; Score 1054; DB 24; Length 4414;
Matches 1075; Conservative 96.8%; Pred. No. 3.2e-282;
0; Mismatches 35; Indels 0; Gaps 0;

QY 1 TCCTTCACAGCATTTGGAAAGCTCCTGAAGACAGTCAATTTGGCAGTTTCATGTGAGAG 60
Db 1717 TCCTTCACAGCATTTGGAAAGCTCCTGAAGACAGTCAATTTGGCAGTTTCATGTGAGAG 60
QY 61 TGAAGTAAACCTGGACCGTTATCAACACAGCTTTAGAAGAGTATATCGTGGCTTCTTTC 120
Db 1777 TGAAGTAAACCTGGACCGTTATCAACACAGCTTTAGAAGAGTATATCGTGGCTTCTTTC 120
QY 121 TGCTGAGGACACATTCGACGACAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 180
Db 1837 TGCTGAGGACACATTCGACGACAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 180
QY 181 CCAGTTTCATACATGAGGGGTACATGATGGATTTGACAGCCCATCAGGCGCGGTGG 240
Db 1897 CCAGTTTCATACATGAGGGGTACATGATGGATTTGACAGCCCATCAGGCGCGGTGG 240
QY 241 TAATATTCTACAATTTGGAAAGTAAAGCTGATTTGAACAGGAAATTTATCAGAAAGATGAAGA 2016
Db 1957 TAATATTCTACAATTTGGAAAGTAAAGCTGATTTGAACAGGAAATTTATCAGAAAGATGAAGA 2016
QY 301 AACTGAAGTACAAGAGCAGATGAATCTCTAAATTTCAAGATGGGAATGCCTCAGGGTAGC 360
Db 2017 AACTGAAGTACAAGAGCAGATGAATCTCTAAATTTCAAGATGGGAATGCCTCAGGGTAGC 360
QY 361 TAGCATGGAAAAACAACCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAAACT 420
Db 2077 TAGCATGGAAAAACAACCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAAACT 420
QY 421 GAAAGAGTTGAATGACTGGCTAACAACCAAGAGAAAGAAAGAAATGGAGGAAGA 2136
Db 2137 GAAAGAGTTGAATGACTGGCTAACAACCAAGAGAAAGAAAGAAATGGAGGAAGA 2136
QY 481 GCCTCTTGACCTGATCTTGAAGACCTAAACGCCAAGTACAACACATAGGTGCTTCA 540
Db 2197 GCCTCTTGACCTGATCTTGAAGACCTAAACGCCAAGTACAACACATAGGTGCTTCA 540
QY 541 AGAAGATCTAGAACCAAGAACAGTCAAGGTCAATTTCTCACTCACATGGTGGTAGT 600
Db 2257 AGAAGATCTAGAACCAAGAACAGTCAAGGTCAATTTCTCACTCACATGGTGGTAGT 600
QY 601 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTGAAGAAACAACCTTAAGTATGGG 2316
Db 2317 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTGAAGAAACAACCTTAAGTATGGG 2316
QY 661 AGATCGATGGGCAAAACATCTGTAGATGGACAGAACCGCTGGGTCTTTTACAAGACAT 2376
Db 2377 AGATCGATGGGCAAAACATCTGTAGATGGACAGAACCGCTGGGTCTTTTACAAGACAT 2376

Mon Apr 28 09:28:32 2003

CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is human
CC dystrophin minigene delta3447 containing nucleotides 1-1992 (N-terminus,
CC hinge H1 and rods R1, R2 and R3), 8749-10227 (rod R24, hinge H4 and
CC CR domain) and 11047-11058 (dystrophin last 3 amino acids).
XX
SQ Sequence 3446 BP; 1074 A; 766 C; 787 G; 819 T; 0 other;

Query Match 51.1%; Score 1042; DB 24; Length 3446;
Best Local Similarity 96.8%; Pred. No. 6e-279; Indels 1; Gaps 1;
Matches 1074; Conservative 0; Mismatches 35;

QY 1 TCCTTACAGCATTTGGAAGCTCCTGAAGCAAGTCAATTTGGCAGTTCATTTGATGGAGAG 60
Db 960 TCCTTACAGCATTTGGAAGCTCCTGAAGCAAGTCAATTTGGCAGTTCATTTGATGGAGAG 1019
QY 61 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGATATTATCGTGGCTTCTTTC 120
Db 1020 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGATATTATCGTGGCTTCTTTC 1079
QY 121 TGCTGAGGACACATTTGCAAGCACAAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 180
Db 1080 TGCTGAGGACACATTTGCAAGCACAAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 1139
QY 181 CCAGTTTTCATCTAGGGGTACATGATGATTTTGACAGCCCATCAGGGCCGGTGG 240
Db 1140 CCAGTTTTCATCTAGGGGTACATGATGATTTTGACAGCCCATCAGGGCCGGTGG 1199
QY 241 TAATATTCTACAATTTGGGAAGTAAGCTGATTGGAACAGGAAATTTATCAGAAGATGAAGA 300
Db 1200 TAATATTCTACAATTTGGGAAGTAAGCTGATTGGAACAGGAAATTTATCAGAAGATGAAGA 1259
QY 301 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGATGGGAATGCCCTCAGGGTAGC 1319
Db 1260 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGATGGGAATGCCCTCAGGGTAGC 420
QY 361 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAAACT 1378
Db 1320 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATC-GAAACT 480
QY 421 GAAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1438
Db 1379 GAAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 540
QY 481 GCCTCTGGACCTGATCTTGAAGACCTAAACGCCCAAGTACAAACATAGGTGCTTCA 1498
Db 1439 GCCTCTGGACCTGATCTTGAAGACCTAAACGCCCAAGTACAAACATAGGTGCTTCA 600
QY 541 AGAAGATCTAGAACAAAGCAAGTCAAGGTCAATTTCTCTACTCAGATGGTGGTGGTAGT 1558
Db 1499 AGAAGATCTAGAACAAAGCAAGTCAAGGTCAATTTCTCTACTCAGATGGTGGTGGTAGT 660
QY 601 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAAGAACAACTTAAGGTATTGGG 1618
Db 1559 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAAGAACAACTTAAGGTATTGGG 720
QY 661 AGATCGATGGGCAACATCTGTAGATGGACAGAGACCGCTGGTCTTTTACAAAGACAT 1678
Db 1619 AGATCGATGGGCAACATCTGTAGATGGACAGAGACCGCTGGTCTTTTACAAAGACAT 780
QY 721 CCTTCTCAAAATGGCAACGCTTACTGAAGAACAGTCCCTTTTAGTGCATGGCTTTCAGA 1738
Db 1679 CCTTCTCAAAATGGCAACGCTTACTGAAGAACAGTCCCTTTTAGTGCATGGCTTTCAGA 840
QY 781 AAAAGAAGATGCAGTGAACAAGATTCACACAACCTGGCTTTAAAGATCAAAATGAAATGTT 1798
Db 1739 AAAAGAAGATGCAGTGAACAAGATTCACACAACCTGGCTTTAAAGATCAAAATGAAATGTT 900
QY 841 ATCAAGTCTTCAAAACTGGCCGTTTAAAGCGGATCTAGAAAAAGAAAAAGCAATCCAT 1858
Db 1799 ATCAAGTCTTCAAAACTGGCCGTTTAAAGCGGATCTAGAAAAAGAAAAAGCAATCCAT

QY 721 CCTTCTCAAAATGGCAACGCTTACTGAAGAACAGTCCCTTTTAGTGCATGGCTTTCAGA 780
Db 2437 CCTTCTCAAAATGGCAACGCTTACTGAAGAACAGTCCCTTTTAGTGCATGGCTTTCAGA 2496
QY 781 AAAAGAAGATGCAGTGAACAAGATTCACACAACCTGGCTTTAAAGATCAAAATGAAATGTT 840
Db 2497 AAAAGAAGATGCAGTGAACAAGATTCACACAACCTGGCTTTAAAGATCAAAATGAAATGTT 2556
QY 841 ATCAAGTCTTCAAAACTGGCCGTTTAAAGCGGATCTAGAAAAAGAAAAAGCAATCCAT 2616
Db 2557 ATCAAGTCTTCAAAACTGGCCGTTTAAAGCGGATCTAGAAAAAGAAAAAGCAATCCAT 2676
QY 901 GGGCAAACTGTATTCACTCAAAACAGATCTTTTCAACACTGAAGAATAAGTCAGTGAC 960
Db 2617 GGGCAAACTGTATTCACTCAAAACAGATCTTTTCAACACTGAAGAATAAGTCAGTGAC 1020
QY 961 CCAGAAGACGGGAGCATGGCTGGATAACTTTGCCCGGTGTTGGGATAAATTTAGTCCAAA 2736
Db 2677 CCAGAAGACGGGAGCATGGCTGGATAACTTTGCCCGGTGTTGGGATAAATTTAGTCCAAA 1080
QY 1021 ACTTGAAGAGAGTACAGCAGACACTCATAGATTACTGCAACAGTTCCTCCCTGGACCTGA 2796
Db 2737 ACTTGAAGAGAGTACAGCAGACACTCATAGATTACTGCAACAGTTCCTCCCTGGACCTGA 1110
QY 1081 AAAGTTTCTTGCCTGGCTTACAGACCTGA 1110
Db 2797 TGAGCTGGACCTCAAGCTGGCCCAAGCTGA 2826

RESULT 9
AAD37242
ID AAD37242 standard; DNA; 3446 BP.
XX AC AAD37242;
XX DT 21-AUG-2002 (first entry)
XX DE Human dystrophin minigene delta3447.
XX KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX OS Homo sapiens.
XX PN WO200183695-A2.
XX PD 08-NOV-2001.
XX PF 27-APR-2001; 2001WO-US13677.
XX PR 28-APR-2000; 2000US-200777P.
XX PA (XIAO/) XIAO X.
XX PI Xiao X;
XX WPI; 2002-049342/06.
XX DR New dystrophin minigene for treating Duchenne or Becker muscular
XX PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
XX PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
XX PT dystrophin gene -
XX PS Example 1; Page 53-54; 71pp; English.
XX CC The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control

Mon Apr 28 09:28:32 2003

RESULT 11
AAD37232 standard; DNA; 2169 BP.

XX AAD37232;

DE 21-AUG-2002 (first entry)

XX Human dystrophin rod, hinge and CR domain regions encoding DNA #1.

XX Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.

XX Homo sapiens.

OS WO200183695-A2.

XX 08-NOV-2001.

XX 27-APR-2001; 2001WO-US13677.

XX 28-APR-2000; 2000US-200777P.

XX (XIAO/) XIAO X.

XX Xiao X;

XX WPI; 2002-049342/06.

DR New dystrophin minigene for treating Duchenne or Becker muscular
XX dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -

XX Example 1; Page 45-46; 71pp; English.

XX The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is human
CC dystrophin DNA fragment encoding rods R22, R23 and R24, hinge H4 and
CC CR domain regions.

XX Sequence 2169 BP; 623 A; 529 C; 524 G; 493 T; 0 other;

SQ Query Match 48.9%; Score 999; DB 24; Length 2169;
Best Local Similarity 100.0%; Pred. No. 4.3e-267;
Matches 999; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1043 ACTCATAGATTACTGCAACAGTTCCCTCCCTGGACCTGGAAAGTTTCTTGCCTGGCTTACA 1102
DB 2 ACTCATAGATTACTGCAACAGTTCCCTCCCTGGACCTGGAAAGTTTCTTGCCTGGCTTACA 61

QY 1103 GAAGCTGAAACAACTGCCAATGTCTACAGGATGCTACCCGTGAAGAAAGGCTCCTAGAA 1162
DB 62 GAAGCTGAAACAACTGCCAATGTCTACAGGATGCTACCCGTGAAGAAAGGCTCCTAGAA 121

QY 1163 GACTCCAAAGGGAGTAAAGAGCTGATGAAACAATGGCAAGACCTCCCAAGGTGAATTGAA 1222
DB 122 GACTCCAAAGGGAGTAAAGAGCTGATGAAACAATGGCAAGACCTCCCAAGGTGAATTGAA 181

QY 1223 GCTCACACAGATGTTTATCACAACTGGATGAAACACAGCCAAAAAATCCTGAGATCCCTG 1282
DB 182 GCTCACACAGATGTTTATCACAACTGGATGAAACACAGCCAAAAAATCCTGAGATCCCTG 241

QY 1283 GAAGGTTCCGATGATGCGAGTCTGTACAAAGACGTTTGGATAACATGAACCTCAAGTGG 1342
DB 242 GAAGGTTCCGATGATGCGAGTCTGTACAAAGACGTTTGGATAACATGAACCTCAAGTGG 301
QY 1343 AGTGAACCTTCGAAAAAGTCTCTCAACATTAGGTCCCAATTTGGAAGCCAGTTCTGACCAG 1402
DB 302 AGTGAACCTTCGAAAAAGTCTCTCAACATTAGGTCCCAATTTGGAAGCCAGTTCTGACCAG 361
QY 1403 TGGAAAGCGTCTGCACCTTTCTTCGAGGAACCTTCTGGTGTGGGTACAGCTGAAAGATGAT 1462
DB 362 TGGAAAGCGTCTGCACCTTTCTTCGAGGAACCTTCTGGTGTGGGTACAGCTGAAAGATGAT 421
QY 1463 GAATTAAGCCGGCAGGCACCTATTGGAGCGGACTTTCCAGCAGTTTCCAGAGCAGAACGAT 1522
DB 422 GAATTAAGCCGGCAGGCACCTATTGGAGCGGACTTTCCAGCAGTTTCCAGAGCAGAACGAT 481
QY 1523 GTACATAGGGCCTTCAAGAGGGAATTGAAAACTAAAGAACCTGTAAATCATGAGTACTCTT 1582
DB 482 GTACATAGGGCCTTCAAGAGGGAATTGAAAACTAAAGAACCTGTAAATCATGAGTACTCTT 541
QY 1583 GAGACTGTACGAATATTCTGACAGAGCAGCCCTTTGGAAGGACTAGAGAACTCTACCAG 1642
DB 542 GAGACTGTACGAATATTCTGACAGAGCAGCCCTTTGGAAGGACTAGAGAACTCTACCAG 601
QY 1643 GAGCCAGAGAGCTGCCTCTGAGGAGAGAGCCCGCAGAAATGTCACTCGGCTTCTACGAAAG 1702
DB 602 GAGCCAGAGAGCTGCCTCTGAGGAGAGAGCCCGCAGAAATGTCACTCGGCTTCTACGAAAG 661
QY 1703 CAGGCTGAGGAGGTCAATACTAGTGGGAAAAATTTGAACCTGCAGCTCCGCTGACTGGCAG 1762
DB 662 CAGGCTGAGGAGGTCAATACTAGTGGGAAAAATTTGAACCTGCAGCTCCGCTGACTGGCAG 721
QY 1763 AGAAAAATAGATGAGACCCCTTGAAGACTCCAGGAACCTTCAAGAGGCGCACGGATGAGCTG 1822
DB 722 AGAAAAATAGATGAGACCCCTTGAAGACTCCAGGAACCTTCAAGAGGCGCACGGATGAGCTG 781
QY 1823 GACCTCAAGCTGGCCCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGGATCTC 1882
DB 782 GACCTCAAGCTGGCCCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGGATCTC 841
QY 1883 CTCATTGACTCTCTCCAAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAATTGCG 1942
DB 842 CTCATTGACTCTCTCCAAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAATTGCG 901
QY 1943 CCTCTGAAAGAGAAACGTGAGCCACGTCAATGACCTTGTCTGCCAGCTTACCACCTTTGGGC 2002
DB 902 CCTCTGAAAGAGAAACGTGAGCCACGTCAATGACCTTGTCTGCCAGCTTACCACCTTTGGGC 961
QY 2003 ATTCAGCTCTCACCGGTATACCTCAGCACTCTTGGGAAGAC 2041
DB 962 ATTCAGCTCTCACCGGTATACCTCAGCACTCTTGGGAAGAC 1000

RESULT 12

AAD37238 standard; DNA; 3531 BP.

XX AAD37238;

XX 21-AUG-2002 (first entry)

XX Human dystrophin minigene delta3531.

XX Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;

XX adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;

XX Becker muscular dystrophy; ds.

XX Homo sapiens.

XX WO200183695-A2.

XX 08-NOV-2001.

N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control element. The dystrophin minigene in operable linkage with an expression control element, in a recombinant adeno-associated virus or retrovirus is useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is AAV vector plasmid construct containing human dystrophin minigenes, a muscle creatine kinase (MCK) promoter and a small polyA signal sequence.

Sequence 4498 BP; 1251 A; 1118 C; 1123 G; 1006 T; 0 other;

Query Match 48.9%; Score 999; DB 24; Length 4498;
Best Local Similarity 100.0%; Pred. No. 6.3e-267;
Matches 999; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1043	ACTCATAGATTACTGCAACAGTTCCCTGGACCTGGAAAGTTTCTTGCCCTGGCTTACA	1102
Db	2108	ACTCATAGATTACTGCAACAGTTCCCTGGACCTGGAAAGTTTCTTGCCCTGGCTTACA	2167
QY	1103	GAAGCTGAAACAACTGCCAATGTCTACAGGATGCTACCCGTAAGGAAAGGCTCCTAGAA	1162
Db	2168	GAAGCTGAAACAACTGCCAATGTCTACAGGATGCTACCCGTAAGGAAAGGCTCCTAGAA	2227
QY	1163	GACTCCAAAGGGAGTAAAGAGCTGATGAAACAATGGCAAGACCTCCAAAGGTGAATGAA	1222
Db	2228	GACTCCAAAGGGAGTAAAGAGCTGATGAAACAATGGCAAGACCTCCAAAGGTGAATGAA	2287
QY	1223	GCTCACACAGATGTTTATCACAACTGGATGAAACACAGCCCAAAAAATCCTGAGATCCCTG	1282
Db	2288	GCTCACACAGATGTTTATCACAACTGGATGAAACACAGCCCAAAAAATCCTGAGATCCCTG	2347
QY	1283	GAAGTCCCGATGATGCAGTCCCTGTACAAAGACGTTTGGATAACATGAACCTTCAAGTGG	1342
Db	2348	GAAGTCCCGATGATGCAGTCCCTGTACAAAGACGTTTGGATAACATGAACCTTCAAGTGG	2407
QY	1343	AGTGAACCTCGGAAAGTCTCTCAACATAGGTCCCAATTTGGAAGCCAGTTCTGACCAG	1402
Db	2408	AGTGAACCTCGGAAAGTCTCTCAACATAGGTCCCAATTTGGAAGCCAGTTCTGACCAG	2467
QY	1403	TGGAAGCGTCTGCACCTTTCTCTGCAGGAAGTCTCTGAGTGGCTACAGCTGAAAGATGAT	1462
Db	2468	TGGAAGCGTCTGCACCTTTCTCTGCAGGAAGTCTCTGAGTGGCTACAGCTGAAAGATGAT	2527
QY	1463	GAATTAAGCCGGCAGCACCTATTGGAGGCGACTTTCCAGCAGTTCCAGAACGAGAACGAT	1522
Db	2528	GAATTAAGCCGGCAGCACCTATTGGAGGCGACTTTCCAGCAGTTCCAGAACGAGAACGAT	2587
QY	1523	GTACATAGGGCCTTCAAGAGGGAATTGAAACCTAAAGAACTTGAAGAACTGTAATCATGACTCTT	1582
Db	2588	GTACATAGGGCCTTCAAGAGGGAATTGAAACCTAAAGAACTTGAAGAACTGTAATCATGACTCTT	2647
QY	1583	GAGACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAAGGAGCTAGAGAACTCTACCAG	1642
Db	2648	GAGACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAAGGAGCTAGAGAACTCTACCAG	2707
QY	1643	GAGCCACAGAGAGCTGCCTCCTGAGGAGAGAGCCGAGATGTCACTCGGCTTCTACGAAAG	1702
Db	2708	GAGCCACAGAGAGCTGCCTCCTGAGGAGAGAGCCGAGATGTCACTCGGCTTCTACGAAAG	2767
QY	1703	CAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTTGAACCTGCACTCCGCTGACTGGCAG	1762
Db	2768	CAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTTGAACCTGCACTCCGCTGACTGGCAG	2827
QY	1763	AGAAAAATAGATGAGACCTTTGAAAGACTCCAGGAACCTTCAAGAGGCGACGGATGAGCTG	1822
Db	2828	AGAAAAATAGATGAGACCTTTGAAAGACTCCAGGAACCTTCAAGAGGCGACGGATGAGCTG	2887
QY	1823	GACCTCAAGCTGCGCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGGATCTC	1882
Db	2888	GACCTCAAGCTGCGCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGGATCTC	2947

QY	1883	CTCATTTGACTCTCTCCAAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAATTCGG	1942
Db	2948	CTCATTTGACTCTCTCCAAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAATTCGG	3007
QY	1943	CCTCTGAAAGAGAACGTGAGCCACGTCAATGACCTTGCTCGCCAGCTTACCACCTTTGGGC	2002
Db	3008	CCTCTGAAAGAGAACGTGAGCCACGTCAATGACCTTGCTCGCCAGCTTACCACCTTTGGGC	3067
QY	2003	ATTGAGCTCTCACCGGTATTAACCTCAGCACTCTGGAAGAC	2041
Db	3068	ATTGAGCTCTCACCGGTATTAACCTCAGCACTCTGGAAGAC	3106

RESULT 14

AAZ48568
ID AAZ48568 standard; cDNA to mRNA; 4402 BP.
XX
AC AAZ48568;

31-MAR-2000 (first entry)

A rod shortened dystrophin (deltaDysAH3) encoding nucleotide sequence.

Muscular dystrophy; rod domain; adeno-associated virus; AAV;
dystrophin gene; truncated; ds.

Homo sapiens.

JP11318467-A.

24-NOV-1999.

08-MAY-1998; 98JP-0142134.

08-MAY-1998; 98JP-0142134.

(KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.

(KOKU-) KOKURITSU SEISHIN SHINKAI CENT SOCHO.

WPI; 2000-100771/09.

P-PSDB; AAY59239.

A gene encoding a shortened dystrophin - useful for the treatment of muscular dystrophy

Claim 7; Page 21-22; 44pp; Japanese.

The invention provides a gene for the treatment of muscular dystrophy having at least one rod repeat structure of hinge 1, hinge 4 and rod domain of dystrophin gene and having a base sequence of 4.5 kb. The gene and a gene-introducing medium consisting of an adeno-associated virus (AAV) vector or lentivirus vector containing the rod shortened dystrophin genes can be used for the genetic treatment of muscular dystrophy of low immune reaction. The present sequence represents a rod shortened dystrophin encoding sequence.

Sequence 4402 BP; 1339 A; 984 C; 1010 G; 1069 T; 0 other;

Query Match 41.7%; Score 851.6; DB 21; Length 4402;
Best Local Similarity 99.5%; Pred. No. 6.2e-226;
Matches 854; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY	1	TCCTTCACAGCATTTGGAAGCTCCTGAGACAGCTTATGAGGAGTATTCGTCGCTTCTTC	60
Db	1155	TCCTTCACAGCATTTGGAAGCTCCTGAGACAGCTTATGAGGAGTATTCGTCGCTTCTTC	1214

QY	61	TGAAGTAAACCTGGACCGTTATCAACACAGCTTTAGAGGAGTATTCGTCGCTTCTTC	120
Db	1215	TGAAGTAAACCTGGACCGTTATCAACACAGCTTTAGAGGAGTATTCGTCGCTTCTTC	1274

QY	121	TGCTGAGGACACATTCGACAGCAGGAGAGATTTCTAATGATGTGGAAGTGTGAAAGA	180
Db	1275	TGCTGAGGACACATTCGACAGCAGGAGAGATTTCTAATGATGTGGAAGTGTGAAAGA	1334

13-JUL-1989.

0.

us-09-845-416-2_copy_960_3000.rng

Mon Apr 28 09:28:32 2003

Db 662 TTATGAATCCAGCGGTGATCATGCAACAGCTGCTTTGGAAGAACAACTTAAGGTACTGGG 721
QY 661 AGATCGATGGGCAACACATCTGTAGATGGACAGAACCCGCTGGGTTCTTTTACAAGACAT 720
Db 722 AGATCGATGGGCAAAATATCTGCAGATGGACTGAAGACCCGGTGGATTGTTTACAAGATAT 781
QY 721 CCTTCTCAAAATGGCAACGCTTTACTGAAGAACAGTGCCTTTTAAAGTGCATGGCTTTCAGA 780
Db 782 TCTTCTAAATGGCAGCATTTTACTGAAGAACAGTGCCTTTTAAAGTGCATGGCTTTCAGA 841
QY 781 AAAAGAAGATGCAGTGAACAAGATTACACAACTGGCTTTTAAAGATCAAAATGAAATGTT 840
Db 842 AAAAGAAGATGCATGAAGACATTCAGACAAAGTGGCTTTTAAAGATCAAAATGAAATGAT 901
QY 841 ATCAAGTCTTCAAAACTGGCCGTTTAAAGCGGATCTAGAAAAGAAAAGCAATCCAT 900
Db 902 GTCAAGTCTTCACAAATATCTACTTTAAATAATAGATCTAGAAAAGAAAAGCAATCCAT 961
QY 901 GGGCAAACTGTATTCACTCAAAACAAGATCTTCTTCAACACTGAAGAATAAGTCAAGTAC 960
Db 962 GGAATACTAAGTTCACCTCAATCAAGATCTACTTTCCGCACTGAAAATAAGTCAAGTAC 1021
QY 961 CCAGAAGACGGAAGCATGGCTGGATAAATTTGCCCGGTGTTGGGATAATTTAGTCCAAAA 1020
Db 1022 TCAAAAGATGGAATCTGGATGGAAAACCTTTGCACAACGTTGGGACAATTTAACCCAAAA 1081
QY 1021 ACTTGAAAAGAGTACAGCACAGACT 1045
Db 1082 ACTTGAAAAGAGTTCAGCACAAAATT 1106

Search completed: April 25, 2003, 08:39:12
Job time : 364.763 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 07:58:12 ; Search time 1925.12 Seconds
(without alignments)
17170.339 Million cell updates/sec

Title: US-09-845-416-2_COPY_960_3000
Perfect score: 2041
Sequence: 1 tcttcacagcatttgaag.....acctcagcactctggaagac 2041

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308016

Minimum DB seq length: 0
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: gb_gss:*
- 18: em_gss_hum:*
- 19: em_gss_inv:*
- 20: em_gss_pln:*
- 21: em_gss_vrt:*
- 22: em_gss_fun:*
- 23: em_gss_mam:*
- 24: em_gss_mus:*
- 25: em_gss_other:*
- 26: em_gss_pro:*
- 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	592.4	29.0	834	13	BI729851
2	491.6	24.1	750	13	BI730168
3	486.6	23.8	579	9	AL121550
4	454.6	22.3	1490	11	BC009242
5	424.2	20.8	663	13	BM488464
6	343	16.8	2334	11	BC011062

7	265.8	13.0	772	13	BI250598
8	226	11.1	250	12	BF963618
C 9	182.4	8.9	645	10	AW467977
C 10	177	8.7	646	13	BI289102
11	172.8	8.5	835	13	BI553820
12	165.2	8.1	600	13	BI988528
C 13	162	7.9	427	12	BE817979
14	153.2	7.5	885	11	AK020881
C 15	150.6	7.4	515	17	AZ780914
16	144.8	7.1	434	14	BQ375536
17	142.2	7.0	426	14	BQ319056
18	142.2	7.0	481	10	BE557463
C 19	140.6	6.9	649	14	BQ209933
20	140	6.9	515	10	BB651473
21	133.2	6.5	521	17	AZ226551
C 22	131.6	6.4	410	14	BQ349936
23	131.6	6.4	784	12	BG212445
C 24	125.6	6.2	504	13	BI289382
25	122.4	6.0	358	12	BF772373
26	118.6	5.8	346	12	BF772375
27	115.6	5.7	449	12	BF522118
28	113.8	5.6	697	10	AV704460
29	103.2	5.1	317	12	BE815203
30	102.6	5.0	312	10	AV308105
31	97.4	4.8	555	9	AI789349
32	96.2	4.7	696	13	BJ075057
C 33	95.4	4.7	297	9	AV297660
34	91.6	4.5	386	9	AI713038
35	86.4	4.2	267	14	BQ379136
C 36	85.8	4.2	255	9	AV294966
C 37	83.6	4.1	555	10	AW494075
C 38	81.2	4.0	501	13	BI061271
C 39	81	4.0	581	13	BM204473
40	79.4	3.9	598	12	BG077459
41	78	3.8	772	13	BM291527
C 42	77.2	3.8	589	17	AQ260288
43	76	3.7	663	10	BB083824
44	73.2	3.6	399	12	BG636373
C 45	72	3.5	332	9	AI715599

ALIGNMENTS

RESULT 1
BI729851
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

BI729851
603349511F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5357162 5',
mRNA sequence.
BI729851
BI729851.1 GI:15706864
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 834)
NIH-MGC <http://mgc.ncbi.nlm.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing Arrayed by: The I.M.A.G.E. Consortium (LLNL)
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM11908 row: e column: 03
High quality sequence stop: 796.
Location/Qualifiers
1. .834

FEATURES
SOURCE

/db_xref="taxon:9606"
/clone="IMAGE:3029414"
/tissue_type="Muscle, rhabdomyosarcoma"
/clone_lib="NIH_MGC_17"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

BASE COUNT 505 a 299 c 328 g 358 t

Query Match 22.3%; Score 454.6; DB 11; Length 1490;
Best Local Similarity 99.1%; Pred. No. 3.6e-97;
Matches 457; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAGTCAATTTGGCAGTTTCATTTGATGGAGAG 60
DB 1029 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAGTCAATTTGGCAGTTTCATTTGATGGAGAG 1088
QY 61 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAGTATTATCGTGGCTTCTTTC 120
DB 1089 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAGTATTATCGTGGCTTCTTTC 1148
QY 121 TGCTGAGGACACATTCGAAGCACAAAGGAGAGATTTCTAATGATGTGGAAAGTGGTGAAGA 180
DB 1149 TGCTGAGGACACATTCGAAGCACAAAGGAGAGATTTCTAATGATGTGGAAAGTGGTGAAGA 1208
QY 181 CCAGTTTCATCTCATGAGGGGTACATGATGGATTTGACAGCCCATCAGGGCCGGTGG 240
DB 1209 CCAGTTTCATCTCATGAGGGGTACATGATGGATTTGACAGCCCATCAGGGCCGGTGG 1268
QY 241 TAATATTCTACAATTGGGAAGTAAGCTGATTTGGAACAGGAGAAATATCAGAAGATGAAGA 300
DB 1269 TAATATTCTACAATTGGGAAGTAAGCTGATTTGGAACAGGAGAAATATCAGAAGATGAAGA 1328
QY 301 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCGAAGATGGGATGCTCAGGGTAGC 360
DB 1329 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCGAAGATGGGATGCTCAGGGTAGC 1388
QY 361 TAGCATGGAANAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGATCAGAAACT 420
DB 1389 TAGCATGGAANAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGATCAGAAACT 1448
QY 421 GAAAGAGTTGAATGACTGGCTAACAACAAACAGAGAAAGAA 461
DB 1449 GAAAGAGTTGAATGACTGGCTAACAACAAACAGAGAAAGAA 1489

RESULT 5
BM488464
LOCUS
DEFINITION
pkm2n.pk007.117 Normalized Chicken Breast Muscle, Leg Muscle, and
Epiphyseal Growth Plate cDNA library (pkm2n) Gallus gallus cDNA
clone pkm2n.pk007.117 5' similar to sp|P11533|DMD-CHICK Dystrophin
pir|S02041 dystrophin, muscle - chicken emb|CAA31746.1| (X13369)
dystrophin (AA 1 - 3660) [Gallus gallus], mRNA sequence.

ACCESSION BM488464
VERSION BM488464.1 GI:18609395
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
Phasianinae; Gallus.

REFERENCE 1 (bases 1 to 663)
AUTHORS Cogburn, L.A. and Monson-Orran, E.
TITLES ESTs from Normalized Chicken Breast Muscle, Leg Muscle, and
Epiphyseal Growth Plate cDNA library, USDA/IFAFS Animal Genome
Project

JOURNAL Unpublished (2002)
COMMENT Contact: Larry A. Cogburn
University of Delaware
Townsend Hall, Newark, DE 19717, USA
Tel: 302-831-1335
Fax: 302-831-2822

Email: cogburn@udel.edu, www.chickest.udel.edu.
Location/Qualifiers
1. .663
/organism="Gallus gallus"
/strains="Commercial broiler and Ottawa Res. Centre
Strains 90 & 21"
/db_xref="taxon:9031"
/clone="pkm2n.pk007.117"
/clone_lib="Normalized Chicken Breast Muscle, Leg Muscle,
and Epiphyseal Growth Plate cDNA library (pkm2n)"
/sex="Male and Female"
/tissue_type="Breast muscle, leg muscle and epiphyseal
growth plate"
/dev_stage="Breast, leg: Embryo (dl9); post-hatch (1d, 1, 3, 5, 7, 9,
11 weeks); growth plate (1d, 7d, 14d post-hatch)"
/lab_host="E. coli EMDH10B"
/note="Vector: pCMVSPORT6; Library made from equivalent
pools of total RNA isolated from each tissue (embryonic
muscle 33.3%; juvenile muscle 33.3%; and epiphyseal growth
plate 33.3% of the final RNA pool). Single pass sequencing
from 5'-end"

BASE COUNT 209 a 132 c 172 g 144 t 6 others
ORIGIN

Query Match 20.8%; Score 424.2; DB 13; Length 663;
Best Local Similarity 78.1%; Pred. No. 4.8e-90;
Matches 518; Conservative 0; Mismatches 144; Indels 1; Gaps 1;

QY 1183 GCTGATGAAACAATGGCAAGACCTCCAAAGGTGAAATTTGAAGTCTCACACAGATGTTTATCA 1242
DB 1 GCTCATGAAGCAGTGGCAGGATCTACAGGCAGAAATTTGATGCACATCTGACATCT-TNN 59
QY 1243 CAACCTGGATGAAACACAGCCAAAATCCTGAGATCCCTGGAGGTTCCTGATGTCAGT 1302
DB 60 CAACCTGGATGAAACACAGCCAAAATCCTGAGATCCCTGGAGGTTCCTGATGTCAGT 119
QY 1303 CCTGTTACAAAGACGTTTGGATAACATGAACTTCAAGTGGAGTGAACCTTCGAAAAAGTC 1362
DB 120 CCTGTTGACAGAGACGTTGGATAACATGAACTTCAAGTGGAGTGAACCTTCGAAAAAGTC 179
QY 1363 TCTCAACATTAGTCCCATTGGAAGCCAGTTCTGACAGTGGAGCGCTGCACCTTTC 1422
DB 180 TCTAAACATTAGATCTCATTTGGAAGCCAGCACAGACCGTGGAGCGCTTTACATCTCTC 239
QY 1423 TCTGCAGGAACCTTCTGGTGTGGCTACAGCTGAAGATGATGAATTAAGCCGGCAGGCACC 1482
DB 240 TCTTCAGGAACCTTTGGCATGGTGCATTTGGAAGCCAGCACAGACCGTGGAGCGCTTTC 299
QY 1483 TATTGGAGGCGACTTTCCAGCAGTTTCAGAAAGCAGAACGATGTACATAGGGCTTCAAGAG 1542
DB 300 CATTGGTGGAGATATTTCCACTGTGCAGAAAGCAGAAATGATGTTTCATAGGACTTTCAAGAG 359
QY 1543 GGAATTGAAAACTAAAGAACCTGTAATCATGATGATCTCTTGAGACTGTACGAATATTTCT 1602
DB 360 GGAGCTGAAAAACAAAGAACCTGTTATCATGAATGCACCTTGAGACTGTGCGACTCTTCTCT 419
QY 1603 GACAGAGCAGCCCTTTGGAAGGACTAGAAAACTCTACCCAGGAGCCAGAGAGCTGCCTCC 1662
DB 420 GGCAGATCAACCCAGTAGAGGACTGGAAAGGTCTATCCAGAAACCAAGAGACCTATCACC 479
QY 1663 TGAGGAGAGAGCCAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATAC 1722
DB 480 TGAGGAGAGGCCCCAGAAATGTCACTAAAGTTCTCCGAAAGGCAAGCAGATGATGTCAAGNAC 539
QY 1723 TGAGTGGAAAAAATTTGAACCTGCCTCCGCTGACTGGCAGAGAAAAATAGATGAGACCTT 1782
DB 540 TGAGTGGGATAAGCTAAATCTACGTTCTGCTGATTGGCAAAAGAGATGATGATGCTCT 599
QY 1783 TGAAGACTCCAGGAACCTTCAAGAGGCGCACGGATGAGCTGGACCTCAAGCTGCGCCAGGC 1842
DB 600 TGAAGACTGCAGGGTCTTTCAGGAGGCGCAATGATGAACTNNNCCTGAAACTGCGCCAGGC 659
QY 1843 TGA 1845

Db 660 TGA 662

RESULT 6

BC011062

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

COMMENT

BC011062
Mus musculus, Similar to utrophin, clone IMAGE:3979320, mRNA.
BC011062.1 GI:15029695
HTC.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 2334)
Strausberg, R.
Direct Submission
Submitted (25-JUL-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
CDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 23 Row: j Column: 4
This clone was selected for full length sequencing because it
passed the following selection criteria: Hexamer frequency ORF
analysis, Similarity but not identity to protein
This clone has the following problem: incomplete processing.
Location/Qualifiers
1. 2334
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="FVB/N"
/clone="IMAGE:3979320"
/tissue_type="Mammary tumor. Metallothionien-TGF alpha
model. 10 month old virgin mouse. Taken by biopsy."
/clone_lib="NCI_CGAP_Mam1"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6"
BASE COUNT 698 a 512 c 638 g 486 t
ORIGIN

Query Match
Best Local Similarity 16.8%; Score 343; DB 11; Length 2334;
Matches 589; Conservative 0; Mismatches 410; Indels 0; Gaps 0;
QY 44 AGTTCATTGATGAGAGTGAAGTAACTGGACCGTTATCAACACGCTTTAGAGAAGTA 103
Db 1042 AGCACCCTGACTGAAGTGGACATGGATTTGGACAGCTACCAGATAGCGCTAGAGGAAGTG 1101
QY 104 TTATCGTGGCTCTTTCTGCTGAGGACACATTGCAAGCACAGGAGAGATTTCTAATGAT 163
Db 1102 CTGACGTGGCTGCTGCCGCGAGGACACGTTCCAGGAGCAAGATGACATTTCTGTATGAT 1161
QY 164 GTGGAAGTGGTGAAGACACAGTTTCATCTCATGAGGGGTACATGATGGATTGACAGCC 223
b 1162 GTCCAAGAAGTCAAGAGCAGTTTGTCTACCCATGAAACTTTTATGATGGAGCTGACAGCA 1221

FEATURES
source
1. 2334
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="FVB/N"
/clone="IMAGE:3979320"
/tissue_type="Mammary tumor. Metallothionien-TGF alpha
model. 10 month old virgin mouse. Taken by biopsy."
/clone_lib="NCI_CGAP_Mam1"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6"
BASE COUNT 698 a 512 c 638 g 486 t
ORIGIN

Query Match
Best Local Similarity 16.8%; Score 343; DB 11; Length 2334;
Matches 589; Conservative 0; Mismatches 410; Indels 0; Gaps 0;
QY 44 AGTTCATTGATGAGAGTGAAGTAACTGGACCGTTATCAACACGCTTTAGAGAAGTA 103
Db 1042 AGCACCCTGACTGAAGTGGACATGGATTTGGACAGCTACCAGATAGCGCTAGAGGAAGTG 1101
QY 104 TTATCGTGGCTCTTTCTGCTGAGGACACATTGCAAGCACAGGAGAGATTTCTAATGAT 163
Db 1102 CTGACGTGGCTGCTGCCGCGAGGACACGTTCCAGGAGCAAGATGACATTTCTGTATGAT 1161
QY 164 GTGGAAGTGGTGAAGACACAGTTTCATCTCATGAGGGGTACATGATGGATTGACAGCC 223
b 1162 GTCCAAGAAGTCAAGAGCAGTTTGTCTACCCATGAAACTTTTATGATGGAGCTGACAGCA 1221

Query Match
Best Local Similarity 16.8%; Score 343; DB 11; Length 2334;
Matches 589; Conservative 0; Mismatches 410; Indels 0; Gaps 0;
QY 44 AGTTCATTGATGAGAGTGAAGTAACTGGACCGTTATCAACACGCTTTAGAGAAGTA 103
Db 1042 AGCACCCTGACTGAAGTGGACATGGATTTGGACAGCTACCAGATAGCGCTAGAGGAAGTG 1101
QY 104 TTATCGTGGCTCTTTCTGCTGAGGACACATTGCAAGCACAGGAGAGATTTCTAATGAT 163
Db 1102 CTGACGTGGCTGCTGCCGCGAGGACACGTTCCAGGAGCAAGATGACATTTCTGTATGAT 1161
QY 164 GTGGAAGTGGTGAAGACACAGTTTCATCTCATGAGGGGTACATGATGGATTGACAGCC 223
b 1162 GTCCAAGAAGTCAAGAGCAGTTTGTCTACCCATGAAACTTTTATGATGGAGCTGACAGCA 1221

Query Match
Best Local Similarity 16.8%; Score 343; DB 11; Length 2334;
Matches 589; Conservative 0; Mismatches 410; Indels 0; Gaps 0;
QY 44 AGTTCATTGATGAGAGTGAAGTAACTGGACCGTTATCAACACGCTTTAGAGAAGTA 103
Db 1042 AGCACCCTGACTGAAGTGGACATGGATTTGGACAGCTACCAGATAGCGCTAGAGGAAGTG 1101
QY 104 TTATCGTGGCTCTTTCTGCTGAGGACACATTGCAAGCACAGGAGAGATTTCTAATGAT 163
Db 1102 CTGACGTGGCTGCTGCCGCGAGGACACGTTCCAGGAGCAAGATGACATTTCTGTATGAT 1161
QY 164 GTGGAAGTGGTGAAGACACAGTTTCATCTCATGAGGGGTACATGATGGATTGACAGCC 223
b 1162 GTCCAAGAAGTCAAGAGCAGTTTGTCTACCCATGAAACTTTTATGATGGAGCTGACAGCA 1221

Query Match
Best Local Similarity 16.8%; Score 343; DB 11; Length 2334;
Matches 589; Conservative 0; Mismatches 410; Indels 0; Gaps 0;
QY 44 AGTTCATTGATGAGAGTGAAGTAACTGGACCGTTATCAACACGCTTTAGAGAAGTA 103
Db 1042 AGCACCCTGACTGAAGTGGACATGGATTTGGACAGCTACCAGATAGCGCTAGAGGAAGTG 1101
QY 104 TTATCGTGGCTCTTTCTGCTGAGGACACATTGCAAGCACAGGAGAGATTTCTAATGAT 163
Db 1102 CTGACGTGGCTGCTGCCGCGAGGACACGTTCCAGGAGCAAGATGACATTTCTGTATGAT 1161
QY 164 GTGGAAGTGGTGAAGACACAGTTTCATCTCATGAGGGGTACATGATGGATTGACAGCC 223
b 1162 GTCCAAGAAGTCAAGAGCAGTTTGTCTACCCATGAAACTTTTATGATGGAGCTGACAGCA 1221

Query Match
Best Local Similarity 16.8%; Score 343; DB 11; Length 2334;
Matches 589; Conservative 0; Mismatches 410; Indels 0; Gaps 0;
QY 44 AGTTCATTGATGAGAGTGAAGTAACTGGACCGTTATCAACACGCTTTAGAGAAGTA 103
Db 1042 AGCACCCTGACTGAAGTGGACATGGATTTGGACAGCTACCAGATAGCGCTAGAGGAAGTG 1101
QY 104 TTATCGTGGCTCTTTCTGCTGAGGACACATTGCAAGCACAGGAGAGATTTCTAATGAT 163
Db 1102 CTGACGTGGCTGCTGCCGCGAGGACACGTTCCAGGAGCAAGATGACATTTCTGTATGAT 1161
QY 164 GTGGAAGTGGTGAAGACACAGTTTCATCTCATGAGGGGTACATGATGGATTGACAGCC 223
b 1162 GTCCAAGAAGTCAAGAGCAGTTTGTCTACCCATGAAACTTTTATGATGGAGCTGACAGCA 1221

Query Match
Best Local Similarity 16.8%; Score 343; DB 11; Length 2334;
Matches 589; Conservative 0; Mismatches 410; Indels 0; Gaps 0;
QY 44 AGTTCATTGATGAGAGTGAAGTAACTGGACCGTTATCAACACGCTTTAGAGAAGTA 103
Db 1042 AGCACCCTGACTGAAGTGGACATGGATTTGGACAGCTACCAGATAGCGCTAGAGGAAGTG 1101
QY 104 TTATCGTGGCTCTTTCTGCTGAGGACACATTGCAAGCACAGGAGAGATTTCTAATGAT 163
Db 1102 CTGACGTGGCTGCTGCCGCGAGGACACGTTCCAGGAGCAAGATGACATTTCTGTATGAT 1161
QY 164 GTGGAAGTGGTGAAGACACAGTTTCATCTCATGAGGGGTACATGATGGATTGACAGCC 223
b 1162 GTCCAAGAAGTCAAGAGCAGTTTGTCTACCCATGAAACTTTTATGATGGAGCTGACAGCA 1221

Query Match
Best Local Similarity 16.8%; Score 343; DB 11; Length 2334;
Matches 589; Conservative 0; Mismatches 410; Indels 0; Gaps 0;
QY 44 AGTTCATTGATGAGAGTGAAGTAACTGGACCGTTATCAACACGCTTTAGAGAAGTA 103
Db 1042 AGCACCCTGACTGAAGTGGACATGGATTTGGACAGCTACCAGATAGCGCTAGAGGAAGTG 1101
QY 104 TTATCGTGGCTCTTTCTGCTGAGGACACATTGCAAGCACAGGAGAGATTTCTAATGAT 163
Db 1102 CTGACGTGGCTGCTGCCGCGAGGACACGTTCCAGGAGCAAGATGACATTTCTGTATGAT 1161
QY 164 GTGGAAGTGGTGAAGACACAGTTTCATCTCATGAGGGGTACATGATGGATTGACAGCC 223
b 1162 GTCCAAGAAGTCAAGAGCAGTTTGTCTACCCATGAAACTTTTATGATGGAGCTGACAGCA 1221

QY 224 CATCAGGCGCGGTTGGTAAATATCTTACAATTTGGAAGTAAGCTGATTTGGAACAGGAAA 283
Db 1222 CACCAGAGCAGCGTGGGGAGCGTCTGCAGGCTGGCAACACAGCTGATGACACAAGGGACT 1281
QY 284 TTATCAGAAGATGAAGAACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGATGG 343
Db 1282 CTGTCAAGAGGAGGAGGAGTGTGAGATCCAGGAACAGATGACCTTGTGAATGCAAGGTG 1341
QY 344 GAATGCCTCAGGCTAGCTAGCATGGAAAAACAAGCAATTTACATAGAGTTTAAATGGAT 403
Db 1342 GAGGCGCTCCGGTGGAGAGCATGGAGAGCGAGTCCCGGCTGCACGACGCTCTGATGGAG 1401
QY 404 CTCCAGATCAGAAACTGAAAGAGTTGAATGACTGGCTAACAACAAAAACAGAAAGAACAA 463
Db 1402 CTGCAGAAAGAAACAGCTGCAGCAGCTCTCAAGCTGGCTGGCCCTCACAGAAGAGCGCAT 1461
QY 464 AGGAAATGGAGGAAGAGCCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCCAAGTACAA 523
Db 1462 CAGAAGATGGAGAGCCCTCCCGCTGGGTGATGACCTGCCCTCCCTGCAGAAAGCTGCTTCAA 1521
QY 524 CAACATAAGGTGCTTCAAGAAAGATCTAGAACAAAGAACTCAGGCTCAATTTCTCTCACT 583
Db 1522 GAACATAAAAGTTTGCATAAATGACCTTGAAGCTGAACAGGTGAAGTAAATTCCTTAAC 1581
QY 584 CACATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 643
Db 1582 CACATGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGG 1641
QY 644 CAACCTTAAGCTATTTGGAGATCGATGGGGGAAACATCTGTAGATGGACAGAGACCGTGG 703
Db 1642 CAGTTACAGAAACTGGGTGAGCGCTGGACAGCTGTATGCCGCTGGACTGAAGAACGTTGG 1701
QY 704 GTTCTTTTACAAGACATCTTCTCAATGGCAACGCTTACTGAAGAAACAGTGCCTTTT 763
Db 1702 AACAGTTGCAAGAAATCAGTATCTGTGGCAGGAATTTATTGAAGAGCAGTGTCTGTG 1761
QY 764 AGTGCATGGCTTTTCAGAAAAAGAGATGCAGTGAACAAGATTACACAACTGGCTTTTAA 823
Db 1762 GAGGCTTGGCTCACCAGAAAGAGAGGCTTTGAATAAAGTTCAAAACCAAGCACTTTAA 1821
QY 824 GATCAAAATGAAATGTATCAAGTCTTCAAAACACTGGCGCTTTTAAAGCGGATCTAGAA 883
Db 1822 GACCAGAGGAACCTAAGTGTCAAGTGTCCCGGCTGTGGCTATATTGAAGGAAGACATG 1881
QY 884 AAGAAAAAGCAATCCATCCATGGCAAACTGTATTCACTCAAAACAAGATCTTCTTCAACACTG 943
Db 1882 ATGAAGAGGCAGACTCTGGATCAACTGATGATGAGTGGCCAGGATGTGGGCCAATTA 1941
QY 944 AAGAATAAGTCAAGTACCCAGAGACGAGAGCATGGCTGGATAACTTTGCCCGGTGTTGG 1003
Db 1942 AGTAATCCCAAGGCATCTAAGAAGATGAACAGTGAAGTGTGAGGAGTAAACACAGAGAT 2001
QY 1004 GATAATTTAGTCCAAAAACTTGAAGAGAGTACAGCACAG 1042
Db 2002 GATTCTCTGGTTCAGAGACTCGAAGACTCTTCTAACCAG 2040

RESULT 7
BI250598
LOCUS
DEFINITION
602993659F1 NCI_CGAP_Mam5 Mus musculus cdna clone linear EST 17-JUL-2001
mRNA sequence.
ACCESSION
BI250598
VERSION
BI250598.1 GI:14799101
KEYWORDS
EST.
SOURCE
house mouse.
ORGANISM
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 772)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)


```
Db 5 GCAACAGTTCCTCCCTGGACCTGGAAAGATTCTTGGCTTGCTTACAGAGCTGAAACAAC 64
QY 1117 TGCCAATGTCTACAGGATGCTACCCGTAAGGAAAGGCTCCTTAGAAGACTCCAAGGGAGT 1176
Db 65 TGCCAATGTCTACAGGATGCTACCCGTAAGGAAAGGCTCCTTAGAAGACTCCAAGGGAGT 124
QY 1177 AAAAGAGCTGATGAAACAATGGCAAGACCTCCAAAGGTGAAATGAAGCTCACACAGATGT 1236
Db 125 AAAAGAGCTGATGAAACAATGGCAAGACCTCCAAAGGTGAAATGAAGCTCACACAGATGT 184
QY 1237 TTATCACACCTGGATGAAACAGCCCAAAATCCTGAGATCCCTG 1282
Db 185 TTATCACACCTGGATGAAACAGCCCAAAATCCTGAGATCCCTG 230

RESULT 9
AW467977/c
LOCUS
DEFINITION
he29g08.x1 NCI_CGAP_CML1 Homo sapiens cDNA clone IMAGE:2920478 3'
similar to SW:UTRO_HUMAN P46939 UTROPHIN ;, mRNA sequence.
ACCESSION
AW467977
VERSION
AW467977.1 GI:7038083
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Elisabeth Paietta, Jonathan D. Licht, M.D.,
Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library Preparation: Life
Technologies, Inc. cDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CCGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 419.
FEATURES
source
Location/Qualifiers
1..645
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2920478"
/clone_lib="NCI_CGAP_CML1"
/tissue_type="myeloid cells, 18 pooled CML cases, BCR/ABL
rearrangement positive, includes both chronic phase and
myeloid blast crisis"
/lab_host="DH10B"
/note="Organ: whole blood; Vector: pCMV-SPORT6; Site_1:
Sall; Site_2: NotI; Cloned unidirectionally. Primer:
Oligo dt. Library constructed by Life Technologies."
BASE COUNT 136 a 162 c 130 g 216 t 1 others
ORIGIN

Query Match
Best Local Similarity 8.9%; Score 182.4; DB 10; Length 645;
Matches 364; Conservative 0; Mismatches 272; Indels 9; Gaps 1;

QY 1082 AAGTTCTTGCCTTACAGACCTGAAACAACACTGCCAATGCTTACAGGATGCTACC 1141
Db 645 AACTCTCTGAAGTGGATCCAGACAGCAGACACAGTGAATGTGCTGTGNATGCTCT 586
QY 1142 CGTAAGGAAAGGCTCCTAGACACTCCAAGGGAGTAAAGAGCTGATGAAACAATGGCAA 1201
Db 585 CATCGGGAGAATGCTCTCAGGATAGTATCTTGGCCAGGGAACCTCAACACAGCAGATGCAG 526
QY 1202 GACCTCCAAGGTGAATTGAAGCTCACACAGATGTTTATCAACAACCTGGATGAAACAACAGC 1261
```

```
Db 525 GACATCCAGGCAGAAATGATGCCCAACAATGACATATTTAAAGCATTTGACGGAACAGG 466
QY 1262 CAAAAAATCCTGAGATCCCTGGAAAGTTCCGATGATGCGAGTCTTGTACAAAGACGTTTG 1321
Db 465 CAGAAGATGGTAAAGCTTTGGGAAATCTGAAGAGGCTACTATGCTTCAACATCGACTG 406
QY 1322 GATAACATGAACCTCAAGTGGAGTGAACCTTCGGAAGAAAGTCTCTCAACATTTAGTCCCAT 1381
Db 405 GATGATATGACCAAGATGAATGACTTAAAGCAAAATCTGCTAGCATCAGGGCCCAT 346
QY 1382 TTGGAAGCCAGTCTGACCAGTGAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTG 1441
Db 345 TTGGAGCCAGCGCTGAGAAAGTGAACAGTGTCTGATGCTCTTAGAAGAACTGATCAAA 286
QY 1442 TGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGCACCTATTGGAGGCGACTTTCCA 1501
Db 285 TGGTGAATATGAAGATGAAGAGCTTAAGAAACAATGCCTATTGGAGAGATGTTCCA 226
QY 1502 GCAGTTTCAAGACAGACGATGACATAGGGCCCTTCAAGAGGGAAATGAAAACTAAAGAA 1561
Db 225 GCCTTACAGCTCCAGTATGACCATGTAAAGCCCTGAGACGGGAGTTAAAGGAGAAAGAA 166
QY 1562 CTTGTAATCATGAGTACTCTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTT--- 1617
Db 165 TATTCTGTCTGAAATGCTGCGACCGAGCCCGAGTTTCTTGGCTGATCAGCCAATTGAG 106
QY 1618 -----GGAAGGACTAGAGAAACTCTACCAGGAGCCCGCAGAGCTGCCTCCTGAGGAGAGA 1672
Db 105 GCCCCTGAAGAGAGCCCAAGAAACCTACAATCAAAACAGAAATTAACCTCCTGAGGAGAGA 46
QY 1673 GCCCAGAATGTCACCTGGCTTCTACGAAAGCAGCGCTGAGGAGGTC 1717
Db 45 GCCCAAAAGATTGCCAAAGCCATGCGCAACACAGTCTTCTGAAGTC 1

RESULT 10
BI289102/c
LOCUS
DEFINITION
UI-R-DK0-cfe-e-07-0-UI.s1 UI-R-DK0 Rattus norvegicus cDNA clone
UI-R-DK0-cfe-e-07-0-UI 3', mRNA sequence.
ACCESSION
BI289102
VERSION
BI289102.1 GI:14946326
KEYWORDS
EST.
SOURCE
Norway rat.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
REFERENCE
1 (bases 1 to 646)
AUTHORS
Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE
Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL
Genome Res. 6 (9), 791-806 (1996)
MEDLINE
97044477
COMMENT
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
The sequence contained an oligo-dt track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dt track served to identify it as a clone from the
normalized rat placenta pool library cDNA Library Preparation: M.B.
Soares Lab Clone distribution: clones will be available through
Research Genetics (www.resgen.com) The following repetitive
elements were found in this cDNA sequence: 1-21,
>AT_rich#Low_complexity
Seq primer: M13 Forward
```


POLYA-Yes.

FEATURES source

Location/Qualifiers

1. .646

/organism="Rattus norvegicus"

/strain="Sprague-Dawley"

/db_xref="taxon:10116"

/clone="UI-R-DK0-cfe-e-07-0-UI"

/clone_lib="UI-R-DK0"

/dev_stage="ADULT"

/lab_host="DH10B (Life Technologies)"

/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DK0 library is a subtracted library derived from a mixture of five individually tagged normalized rat libraries: brain-nRBP (20%), heart-nRHP (20%), kidney-nRKP (20%), aorta-nRAP (20%), and placenta-nRPP (20%). Each original library was constructed from a mixture of equal amounts of RNA from seven different developmental time-points: embryonic day 17, embryonic day 19, embryonic day 21, adult day 1, adult day 12, adult day 75, and adult day 200. (Exception: the aorta pool does not contain embryonic day 17 RNA and the placenta pool contains only the three embryonic stages). Each library was normalized individually according to the procedure described by Bonaldo, Lennon & Soares (Genome Research 6: 791-806, 1996). For construction of the DK0 subtracted library, plasmid DNA from each of the five individually tagged normalized libraries was mixed in the proportions specified above and electroporated into competent bacteria for production of single-stranded circular DNA representing the pool of libraries. Single-stranded circular DNA representing these five normalized libraries was then used as a tracer in a subtractive hybridization with a driver (PCR amplified inserts from a plasmid DNA template preparation) comprising: a) a set of about 1,000 arrayed clones from each of the five non-normalized libraries of brain (CT0s), heart (CS0s), kidney (CU0s), aorta (CW0s), and placenta (CX0s). The resulting pool of approximately 5,000 clones represented about 33.3% of the final driver population. A set of about 2,000 arrayed clones from each of the five normalized libraries of brain (CT0), heart (CS0), kidney (CU0), aorta (CW0), and placenta (CX0). The resulting pool of about 10,000 clones represented about 66.6% of the final driver population.

TAG_LIB=UI-R-DK0

TAG_TISSUE=rat placenta pool

TAG_SEQ=TCACGACAGT

BASE COUNT 131 a 187 c 129 g 198 t 1 others

Query Match 8.7%; Score 177; DB 13; Length 646;

Best Local Similarity 59.6%; Pred. No. 2.5e-31;

Matches 297; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 257 GGAAGTAAGCTGATTGGAACAGGAAATATCAGAAGATGAAGAACTGAAGTACAGAG 316

Db 639 GGCAACCAACTGATGACACAGGCACTCTGTGGATGAGGAGGAGTNTGAGATTCAGGAG 580

QY 317 CAGATGAATCTCCTAAATTCAGATGGGAATGCCTCAGGTACCTAGCATGGAAACAA 376

Db 579 CAGATGACCTTGCTGAACGCCAGATGGAGGCGCTCCGGGTAGACAGCATGGAGAGACAG 520

QY 377 AGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAACTGAAAGAGTTGAATGAC 436

Db 519 TCCCGGTTGCATGATGCCCTGATGGAACCTGCAGAAAGAACAGCTGCAGCAGCTCTCAGGC 460

QY 437 TGGCTAACAAAACAGAGAAAGAAACAGGAAATGGAGGAGAGCCCTCTTGGACCTGAT 496

Db 459 TGGCTGACCCCTCAGAGAGGAGCGCATTCAGAGATGGAAGCCCTCCAGTGGGTGACGAC 400

QY 497 CTTGAAGACCTAAACGCCAAGTACAACAACATAAGGTGCTTCAAGAAGATCTAGAACA 556

Db 399 CTACCCCTCCCTGCRAAACCTTGCTTGAAGAACAATAAAAGTTTGCAAGTGACCTCGAAGCT 340

QY 557 GAACAAGTCAGGTCATCTCTCACTCACATGGTGGTGTAGTTGATGAATCTAGTGA 616

Db 339 GAGCAGGTGAAGGTGAATTCCTTAACATATGGTGGTGTATGTGGATGAAACAGTGGG 280

QY 617 GATCAGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGGAGATCGATGGGCAAC 676

Db 279 GAGAGCCACAGCTGTTTGGAGATCAGTTACAGAAACTGGTGAGCGCTGGACAGCT 220

QY 677 ATCTGTAGATGGACAGAACCGCTGGTCTTTTACAAGACATCTCTCAATGGCAA 736

Db 219 GTATGCCGCTGGACTGAAGAACGTTTGAACAGGTTGCAAGAAATCAATATTTGTGGCAG 160

QY 737 CGTCTTACTGAAGAACAG 754

Db 159 GAATTATTGGAAGAACAG 142

RESULT 11

BI553820 835 bp mRNA linear EST 05-SEP-2001

LOCUS 603190772F1 NIH_MGC_95 Homo sapiens cDNA clone IMAGE:5262209 5',

DEFINITION mRNA sequence.

ACCESSION BI553820

VERSION BI553820.1 GI:15441132

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE

AUTHORS NIH-MGC http://mgc.nci.nih.gov/.

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL Unpublished (1999)

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgapbs-r@mail.nih.gov

Tissue procurement: Miklos Palkovits, M.D., Ph.D.

cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki Toshioyuki and Piero Carninci (RIKEN)

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA sequencing by: Incyte Genomics, Inc.

DNA distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov

Plate: LLAM11660 row: p column: 18

High quality sequence stop: 786.

FEATURES source

Location/Qualifiers

1. .835

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="IMAGE:5262209"

/clone_lib="NIH_MGC_95"

/tissue_type="hippocampus"

/lab_host="DH10B"

/note="Organ: brain; Vector: pBluescriptR (modified pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3', size-selected for average insert size 2.5 kb and normalized to 5. This is a primary library enriched for full-length clones and constructed using the Cap-trapper method (Carninci, in preparation). Library constructed by M. Brownstein (NIH/NHGRI, National Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 220 a 219 c 220 g 176 t

ORIGIN

Query Match 8.5%; Score 172.8; DB 13; Length 835;

Best Local Similarity 56.0%; Pred. No. 2.7e-30;

Matches 327; Conservative 0; Mismatches 257; Indels 0; Gaps 0;

QY 1457 GATGATGAATTAAGCCGCGCAGCACCTATTGGAGCGGACTTTCCAGCAGTTCAGAGCAG 1516

Db 6 GATGTGGAGTTGTGTCAGCTCAGCTGCCCTACAGGGGATGTGGCCCTGGTGTGCAACAGGAG 65

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 08:10:12 ; Search time 46.4852 Seconds
(without alignments)
13465.096 Million cell updates/sec

Title: US-09-845-416-2_COPY_960_3000

Perfect score: 2041
Sequence: 1 tccttcacagcatttgaag.....acctcagcactcttgaagac 2041

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 875442

Minimum DB seq length: 0

Maximum DB seq length: 5000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA:*

1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*

2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*

3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*

4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*

5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*

6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Match	Length	DB	ID	Description
1	40.4	2.0	2223	1	US-08-257-073-4	Sequence 4, Appli
2	38.6	1.9	1995	1	US-08-425-069-3	Sequence 3, Appli
3	38.6	1.9	1995	2	US-08-317-844B-3	Sequence 3, Appli
4	38.2	1.9	1131	6	5180810-3	Patent No. 5180810
5	38.2	1.9	1784	6	5180810-2	Patent No. 5180810
6	38.2	1.9	1848	4	US-09-134-001C-447	Sequence 447, App
7	38.2	1.9	1394	4	US-09-247-155-76	Sequence 76, Appl
8	37.4	1.8	2082	3	US-08-985-335-4	Sequence 4, Appli
9	37.4	1.8	2082	4	US-09-410-372-4	Sequence 4, Appli
10	37.2	1.8	3902	4	US-08-961-527-212	Sequence 212, App
11	36.8	1.8	1886	6	5210183-1	Patent No. 5210183
12	36.6	1.8	1845	4	US-08-887-534A-22	Sequence 22, Appl
13	36.2	1.8	289	4	US-09-007-005-17	Sequence 17, Appl
14	36.2	1.8	289	4	US-09-244-796-17	Sequence 17, Appl
15	36.2	1.8	2447	2	US-09-014-969-14	Sequence 14, Appl
16	36.2	1.8	608	4	US-09-385-982-236	Sequence 236, App
17	36.2	1.8	2873	4	US-08-630-915A-193	Sequence 193, App
18	35.8	1.8	790	4	US-09-461-474-13	Sequence 13, Appl
19	35.6	1.7	4868	1	US-08-139-937-12	Sequence 12, Appl
20	35.6	1.7	4868	5	PCT-US93-11310-12	Sequence 12, Appl
21	35.2	1.7	3466	1	US-08-468-036-38	Sequence 38, Appl
22	35.2	1.7	3466	2	US-08-376-843-38	Sequence 38, Appl
23	35.2	1.7	2169	4	US-09-434-408-3	Sequence 3, Appli
24	35.2	1.7	4766	5	PCT-US93-07261-10	Sequence 10, Appl
25	34.8	1.7	2442	4	US-09-040-485-1	Sequence 1, Appli
26	34.8	1.7	3337	1	US-08-072-610-1	Sequence 1, Appli
27	34.8	1.7	3337	2	US-08-719-822B-1	Sequence 1, Appli

28	34.8	1.7	3337	4	US-09-092-458-1	Sequence 1, Appli
29	34.6	1.7	198	5	PCT-US95-10668-3	Sequence 3, Appli
30	34.6	1.7	198	5	PCT-US95-10668-4	Sequence 4, Appli
31	34.6	1.7	2277	1	US-08-676-967-5	Sequence 5, Appli
32	34.6	1.7	2277	1	US-08-676-974-5	Sequence 5, Appli
33	34.6	1.7	2277	2	US-09-098-487-5	Sequence 5, Appli
34	34.6	1.7	2704	4	US-08-857-076-44	Sequence 44, Appl
35	34.6	1.7	3499	4	US-08-857-076-43	Sequence 43, Appl
36	34.4	1.7	717	4	US-09-461-697-189	Sequence 189, App
37	34.4	1.7	774	4	US-09-461-697-187	Sequence 187, App
38	34.4	1.7	819	4	US-09-461-697-185	Sequence 185, App
39	34.4	1.7	1047	4	US-09-671-950-1	Sequence 1, Appli
40	34.4	1.7	1047	4	US-09-671-950-3	Sequence 3, Appli
41	34.4	1.7	1047	4	US-09-671-950-5	Sequence 5, Appli
42	34.4	1.7	1047	4	US-09-671-950-7	Sequence 7, Appli
43	34.4	1.7	1047	4	US-09-671-950-9	Sequence 9, Appli
44	34.4	1.7	1047	4	US-09-671-950-11	Sequence 11, Appl
45	34.4	1.7	1047	4	US-09-671-950-13	Sequence 13, Appl

ALIGNMENTS

RESULT 1

US-08-257-073-4

; Sequence 4, Application US/08257073

; Patent No. 5766597

; GENERAL INFORMATION:

; APPLICANT: Paoletti, Enzo

; APPLICANT: de Taisne, Charles

; APPLICANT: Tine, John A.

; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE

; NUMBER OF SEQUENCES: 143

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Curtis, Morris & Safford, P.C.

; STREET: 530 Fifth Avenue, 25th Floor

; CITY: New York

; STATE: New York

; COUNTRY: UNITED STATES OF AMERICA

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/257,073

; FILING DATE: 09-JUN-1994

; CLASSIFICATION: 424

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/075,783

; FILING DATE: 11-JUN-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/852,305

; FILING DATE: 18-MAR-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/672,183

; FILING DATE: 20-MAR-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: Frommer, William S.

; REGISTRATION NUMBER: 25,506

; REFERENCE/DOCKET NUMBER: 454310-2570

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 840-3333

; TELEFAX: (212) 840-0712

; TELEX: 425066 CURTMS

; INFORMATION FOR SEQ ID NO: 4:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2223 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

US-08-257-073-4


```

; IMMEDIATE SOURCE:
; CLONE: p6B
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1
US-08-317-844B-3

```

Query Match 1.9%; Score 38.6; DB 2; Length 1995;
Best Local Similarity 44.0%; Pred. No. 0.2;
Matches 164; Conservative 0; Mismatches 209. Indels

		104;	Conservative	0;	Mismatches	209;	Indels	0;	Gaps	0;
QY	1502	GCAGTTCAGAACGACGATGTACATAGGCGCTTCAAGAGGGAATTGAAACTAAAGAA	1561							
Db	731	GCAGTGCAGCTGACGACGCCGACGAGCAGGACCTGGACAACAAGGACCCGGAGGATATGGAC	790							
QY	1562	CCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGAA	1621							
Db	791	CAGGACAAAGGACCATCTGGACCCGGTAGTCCGCTGCAGCAGCAGCGCCGCGCAGCAG	850							
QY	1622	GGACTAGAGAAACTCTACGAGGAGCCCCAGAGAGCTGCCCTCTGAGGAGAGAGCCCCAGAAAT	1681							
Db	851	GACCTGGAGGATATGCCCTTGACACAACAAGGACCCGGAGGATATGGACCAGGACAACAAG	910							
QY	1682	GTCACTCGGCTTCTACGAARACCAGGCTGAGGAGGTCAATACTAGTGGGAAAAATTGAAC	1741							
Db	911	GACCATCTGGAGCAGGCGTGCAGCAGCAGCAGCGCCGAGCAGGACCTGGACAACAAGGAT	970							
QY	1742	CTGCACTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAAGACTCCAGGAACCTT	1801							
Db	971	TAGGAGGTTATGGACCAAGGACAACAAGGTCCAGGAGGATATGGACCAGGACAACAAGGTC	1030							
QY	1802	CAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGATCC	1861							
Db	1031	CAGGAGGATATGGACCAGGTTAGTGCATCTGCAGCAGCAGCCGCGCAGCAGGACCTGGACAAC	1090							
QY	1862	TGGCAGCCCGTGG	1874							
Db	1091	AAGGACCAGGAGG	1103							

RESULT 4
5180810-

```

; Patent No. 5180810
; APPLICANT: Gomi, Hideyuki;Hozumi, Tatsunobu;Hattori, Shizuo;
; Tagawa, Chiaki;Kishimoto, Fumitaka;Bjorck, Lars
; TITLE OF INVENTION: PROTEIN H CAPABLE OF BINDING TO IGG
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/376,641
; FILING DATE: 07-JUL-1989
; SEQ ID NO:3:
; LENGTH: 1131

```

; LENGTH: 1131
5180810-3

```

Query Match          1.9%;      Score 38.2;      DB 6;      Length 1131;
Best Local Similarity 45.9%;
Matches 130; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

```

[illegible]

QY 521 CAACAACATAAAGGTGCTTCAAGAAGATCTAGAACAAAGAACAAG 563
 Db 397 TTAGAACGTAAATACCAACGAGAAGTAGAAAACGTTATCAAG 439

RESULT 5
5180810-2

```

; Patent No. 5180810
; APPLICANT: Gomi, Hideyuki;Hozumi, Tatsunobu;Hattori, Shizuo;
; Tagawa, Chiaki;Kishimoto, Fumitaka;Bjorck, Lars
; TITLE OF INVENTION: PROTEIN H CAPABLE OF BINDING TO IGG
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/376,641
; FILING DATE: 07-JUL-1989
; SEQ ID NO:2:
; LENGTH: 1784
; 5180810-2

```

Query Match 1.9%; Score 38.2; DB 6; Length 1784;
Best Local Similarity 45.9%; Pred. No. 0.24;
Matches 130; Conservative 0; Mismatches 153;

	Conservative	Mismatches	Indels	Gaps
QY 281	AAATTATCAGAAGATGAAGAAAACCTGAAGTACAAGAGCAGATGAATCTCCTAAATTC AAGA	340		
Db 484	AAAAAGTTAGACGAAGATAATGCTAAACTGTTGAGGTGTTGAAACCACAAGTTGGAA	543		
QY 341	TGGSAATGCCCTCAGGGTAGCTAGCATGGAAAAACAAAGCAATTTACATAGAGTTTTAATG	400		
Db 544	AACGAAAAA ACTCAAGAGTGAGAATGAGGAGAATAAGAAAAATTTAGACAAA CTTAGC AAA	603		
QY 401	GATCTCCAGAATCAGAAAAC TGAAGAGTTGAATGACTGGCTTAACAAAAACAGAAGA GA	460		
Db 604	GAAAA TC AAGGAAAGCTCGAAAAAAT TGAGCTTGACATCTCAAAAAAT TAGATCACG AG	663		
QY 461	ACAAGGAAAATGGAGGAAGAGCCCTCTTGGACCTGATCTTGAAGACCTTAAAA CGCCA GTA	520		
Db 664	CACAAAGAGCACCRAAAAAAGAACACAAGAACAAGAGCGCACAAAAAAATCAAGAAC AA	723		
QY 521	CAACAACATAAGGTGCTTCAAGAAAGATCTAGAACAAAGAACAA G	563		
Db 724	TTAGAACGTAATAATCAACACGAGAAGTAGAAAAACGTTATCAAG	766		

RESULT 6

US-09-134-001C-447
; Sequence 447, Application US/09134001C
; Patent No. 6380370
; GENERAL INFORMATION:

APPLICANT: Lynn Dou

```

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC-007
; CURRENT APPLICATION NUMBER: US/09/134,001C
; CURRENT FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/064,964
; PRIOR FILING DATE: 1997-11-08
; PRIOR APPLICATION NUMBER: US 60/055,779
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 5674
; SEQ ID NO 447
; LENGTH: 1848
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-447

```

Query Match

	Query Match	1.9%;	Score 38.2;	DB 4;	Length 1848;
	Best Local Similarity	48.0%;	Pred. NO.	0.25;	
	Matches 109; Conservative	0;	Mismatches 118;	Indels 0;	Gaps 0;
QY	793 AGTGAACAAGATTACACAACTGGCTTTAAAGATCAAAAATGAAATGTTATCAAGTCCTTCA	852			

Mon Apr 28 09:28:33 2003

b 207 AGTTGAAATTTTCTCAATAATCACCTTTTATCGTTAAAGTAGCGGGAAGAAAGTGCCTT 266
y 853 AAAAGTGGCGTTTAAAGCGGATCTAGAAAAGAAAAGCAATCCATGGGCAAACTGTA 912
b 267 AAAAGATACGTTTACTGGTGACATAGTATTAAAAATAAAAAATTGTCTCAATCAAT 326
y 913 TTCACTCAAAACAAGATCTTCTTCAACACACTGAAGATAAAGTCAGTGACCCAGAGACGGA 972
b 327 TGAAGTTTGAAGAACTCTTCAAAATACATACTAACGGAATATTTTCATATCAACATTTCAA 386
y 973 AGCATGGCTGGATAACTTTGCCCGGTGTTGGGATAATTTAGTCCAAA 1019
b 387 ACGATTATACGATGAATTAATACATAGTAGGGATAGTTTGTATGAAA 433

RESULT 7
US-09-247-155-76/c
; Sequence 76, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET.021A
; CURRENT APPLICATION NUMBER: US/09/247,155A
; EARLIER FILING DATE: 1999-02-09
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-08-10
; EARLIER APPLICATION NUMBER: 60/099,273
; EARLIER FILING DATE: 1998-10-04
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patent.pm
; SEQ ID NO 76

LENGTH: 1394
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 46..675
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 46..87
OTHER INFORMATION: Von Heijne matrix
OTHER INFORMATION: score 5.9
OTHER INFORMATION: seq LTLGLSLILAGL/IV
FEATURE:
NAME/KEY: polyA_signal
LOCATION: 1363..1368
FEATURE:
NAME/KEY: polyA_site
LOCATION: 1382..1394
US-09-247-155-76

Query Match 1.9%; Score 38; DB 4; Length 1394;
Best Local Similarity 59.1%; Pred. No. 0.24;
Matches 65; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 369 AAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACTGAAAGAGT 428
Db 1087 AAAACAAATTTAAATATTTTCAAAATTTTAAATTTGCTGCACCATAGATGAATAAGAGC 1028
QY 429 TGAATGACTGGCTAACAAAAACAGAGAAAGAAACAAAGGAAATGGAGGAA 478
Db 1027 TTACTTAAGGAAAGAAAGAAACAAACAAACAAACAAACAAACAAACAAACGATGGA 978

RESULT 8
US-08-985-335-4/c

; Sequence 4, Application US/08985335
; Patent No. 6080847
; GENERAL INFORMATION:
; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Yue, Henry
; APPLICANT: Lal, Preeti
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; TITLE OF INVENTION: PROTEINS ASSOCIATED WITH CELL
; TITLE OF INVENTION: PROLIFERATION
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Dr.
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/985,335
; FILING DATE: Filed Herewith
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Billings, Lucy J.
; REGISTRATION NUMBER: 36,749
; REFERENCE/DOCKET NUMBER: PF-0421 US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-855-0555
; TELEFAX: 650-845-4166
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2082 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; LIBRARY: LATRTUT02
; CLONE: 1352286
; US-08-985-335-4

Query Match 1.8%; Score 37.4; DB 3; Length 2082;
Best Local Similarity 48.4%; Pred. No. 0.47;
Matches 104; Conservative 0; Mismatches 111; Indels 0; Gaps 0;

QY 675 ACATCTGTAGTGGACAGAAGACCGCTGGGTTCTTTTACAAGACATCCTTCTCAATGSC 734
Db 1787 ACAAATGGACAGCGCAGGAGGATGATAGCATTCCTCTGATTACTATATAAGTCAACAATA 1728
QY 735 AACGTCTTACTGAAGAACAGTGCCTTTTGTAGTGCATGGCTTTCAGAAAAGAGATGCAG 794
Db 1727 TACTTATAATTAAGAAAGCTTCCTTTAGCTGGATTACCTGATATAAGGCAACACAG 1668
QY 795 TGAACAAGATTACACACAACTGGCTTTTAAAGATCAAAATGAAATGTTATCAAGTCTTCAA 854
Db 1667 TTAATCAAAATGTAACATATGTTTCTAACACCAGTATAGAAAATACAAACTAATTACA 1608
QY 855 AACTGGCCCGTTTAAAGCGGATCTAGAAAAAGAAA 889
Db 1607 ATCAGAACAAATTTCTACCTTATCTACTATAGTAA 1573

RESULT 9
US-09-410-372-4/c
; Sequence 4, Application US/09410372
; Patent No. 6281334
; GENERAL INFORMATION:

APPLICANT: Hillman, Jennifer L.
 APPLICANT: Yue, Henry
 APPLICANT: Lal, Preeti
 APPLICANT: Shah, Purvi
 APPLICANT: Corley, Neil C.
 TITLE OF INVENTION: PROTEINS ASSOCIATED WITH CELL
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Incyte Pharmaceuticals, Inc.
 STREET: 3174 Porter Dr.
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94304
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq for Windows Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/410,372
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/985,335
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Billings, Lucy J.
 REGISTRATION NUMBER: 36,749
 REFERENCE/DOCKET NUMBER: PF-0421 US
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 650-855-0555
 TELEFAX: 650-845-4166
 INFORMATION FOR SEQ ID NO: 4:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2082 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 IMMEDIATE SOURCE:
 LIBRARY: LATRTUT02
 CLONE: 1352286

US-09-410-372-4
 Query Match 1.8%; Score 37.4; DB 4; Length 2082;
 Best Local Similarity 48.4%; Pred. No. 0.47;
 Matches 104; Conservative 0; Mismatches 111; Indels 0; Gaps 0;
 QY 675 ACATCTGTAGATGACAGAACAGTGCCTTTTAAAGATCAAAATGAAATGATGCGAG 734
 Db 1787 ACAAATGGACAGGCAAGCAGGATGATAGCATCTCTTGATTACTATAATAGTCAACAATA 1728
 QY 735 AACGCTTACTGAAGAACAGTGCCTTTTAAAGATCAAAATGAAATGATGCGAG 794
 Db 1727 TACTTATAATTAAGAAAGCTTCTTTAGCTGGATTACCTGCATAAATAGGCAACACAG 1668
 QY 795 TGAACAAGATTCACACAACTGGCTTTAAAGATCAAAATGAAATGATGCGAGTCTTCAA 854
 Db 1667 TTAATCAATGTAACATATGTTTCTAACACAGTATAGAAAATACAAAATAATTACA 1608
 QY 855 AACTGGCCGTTTAAAGCGGATCTAGAAAAGAA 889
 Db 1607 ATCAGAACAATTTCTACCTTATCTACTACTATAGTAA 1573

RESULT 10
 JS-08-961-527-212/c
 : Sequence 212, Application US/08961527
 : Patent No. 6420135
 : GENERAL INFORMATION:
 APPLICANT: Charles Kunsch
 TITLE OF INVENTION: Streptococcus pneumoniae Polynucleotides and Sequences
 NUMBER OF SEQUENCES: 391

CORRESPONDENCE ADDRESS:
 ADDRESSEE: Human Genome Sciences, Inc.
 STREET: 9410 Key West Avenue
 CITY: Rockville
 STATE: Maryland
 COUNTRY: USA
 ZIP: 20850
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
 COMPUTER: HP Vectra 486/33
 OPERATING SYSTEM: MSDOS version 6.2
 SOFTWARE: ASCII Text
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/961,527
 FILING DATE:
 CLASSIFICATION: 424
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER:
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: Brookes, A. Anders
 REGISTRATION NUMBER: 36,373
 REFERENCE/DOCKET NUMBER: PB340P1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (301) 309-8504
 TELEFAX: (301) 309-8512
 INFORMATION FOR SEQ ID NO: 212:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3902 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 US-08-961-527-212

Query Match 1.8%; Score 37.2; DB 4; Length 3902;
 Best Local Similarity 50.6%; Pred. No. 0.83;
 Matches 90; Conservative 0; Mismatches 88; Indels 0; Gaps 0;
 QY 671 GCAACATCTGTAGATGGACAGACCGCTGGGTTCTTTTACAAGACATCTTCTCAA 730
 Db 2724 GAAAAAGTACGTGTCAAGACTCTATCACTGGTGTGTAGTACCTGATAAATCTACGAA 2665
 QY 731 TGGCAACGTCTTACTGAAGACAGTGCCTTTTGTAGTGCATGGCCTTTCAGAAAAAGAGAT 790
 Db 2664 TTGGCTCGTAAATGAAGAAATGATACCTCTTCAGCCCTTACTCTGTAGAAAAAGATAC 2605
 QY 791 GCAGTGAACAAGATTCACACAACCTGGCTTTAAAGATCAAAATGAAATGATCAAGTC 848
 Db 2604 GGTGTGCCATTCAATATACATGACATTAAGAAAAATACGATGAATGGTCGCAATC 2547

RESULT 11
 5210183-1
 : Patent No. 5210183
 : APPLICANT: LINDAHL, GUNNAR;FRITHZ, ELISABET;HEDEN, LARS-OLOF
 : TITLE OF INVENTION: PROTEIN ARP, WITH IMMUNOGLOBULIN A
 : BINDING ACTIVITY, THE CORRESPONDING VECTORS AND HOSTS, REAGENT
 : KIT AND PHARMACEUTICAL COMPOSITION
 : NUMBER OF SEQUENCES: 3
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/07/270,099
 : FILING DATE: 14-NOV-1988
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 186,097
 : FILING DATE: 25-APR-1988
 : SEQ ID NO:1:
 : LENGTH: 1886
 : 5210183-1

Query Match 1.8%; Score 36.8; DB 6; Length 1886;
 Best Local Similarity 45.5%; Pred. No. 0.68;
 Matches 131; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

Mon Apr 28 09:28:33 2003

274 AACAGGAAATATATCAGAAAGATGAAGAACTGAAGTACAAGAGCAGATGAATCTCTCTAAA 333
261 AGCGTGAAGTGGCTTAAAGAAATATAACGGCTTACTTAAGGAAATGAGGAGCTCAAGGT 320
334 TTCAAGATGGGAATGCCTCAGGGTAGCTAGCATGGAAAAACAAGCAATTTACATAGAGT 393
321 AGAAGTGAAGAAATATCTATCTTATGCTGACGATAAGAAAAAGATCCTCAATCTAGAGC 380
394 TTTAATGGATCTCCAGATCAGAAACTGAAGAGTGAATGACTGGCTAACAACAAACAGA 453
381 ATTAATGGTGAAGAAATCAAGATCTTCGAAAAAGAGAGGAGCAATATCAGGACAAATAGA 440
454 AGAAAGCAACAGGAAATGAGGAGAGAGCCTCTTGGACCTGATCTTGAAGACCTAAAACG 513
441 AGAAGTGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 500
514 CCAAGTACAACAAACATAAGGTGCTTCAAGAAAGATCTAGAACAAGAAACA 561
501 TCAATAGAAAGCAGATAGCATTAATCAAGAAACAACAAAGAAAGAAACATCA 548

RESULT 12
US-08-887-534A-22
; Sequence 22, Application US/08887534A
; Patent No. 6455323
; GENERAL INFORMATION:
; APPLICANT: Holden, David W.
; TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/887,534A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 28341/33996
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: (312) 474-6600
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1845 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA (genomic) (p14c15)"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..788
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 856..1842
US-08-887-534A-22

Query Match 1.8%; Score 36.6; DB 4; Length 1845;
Best Local Similarity 48.0%; Pred. No. 0.77;
Matches 135; Conservative 0; Mismatches 144; Indels 2; Gaps 1;

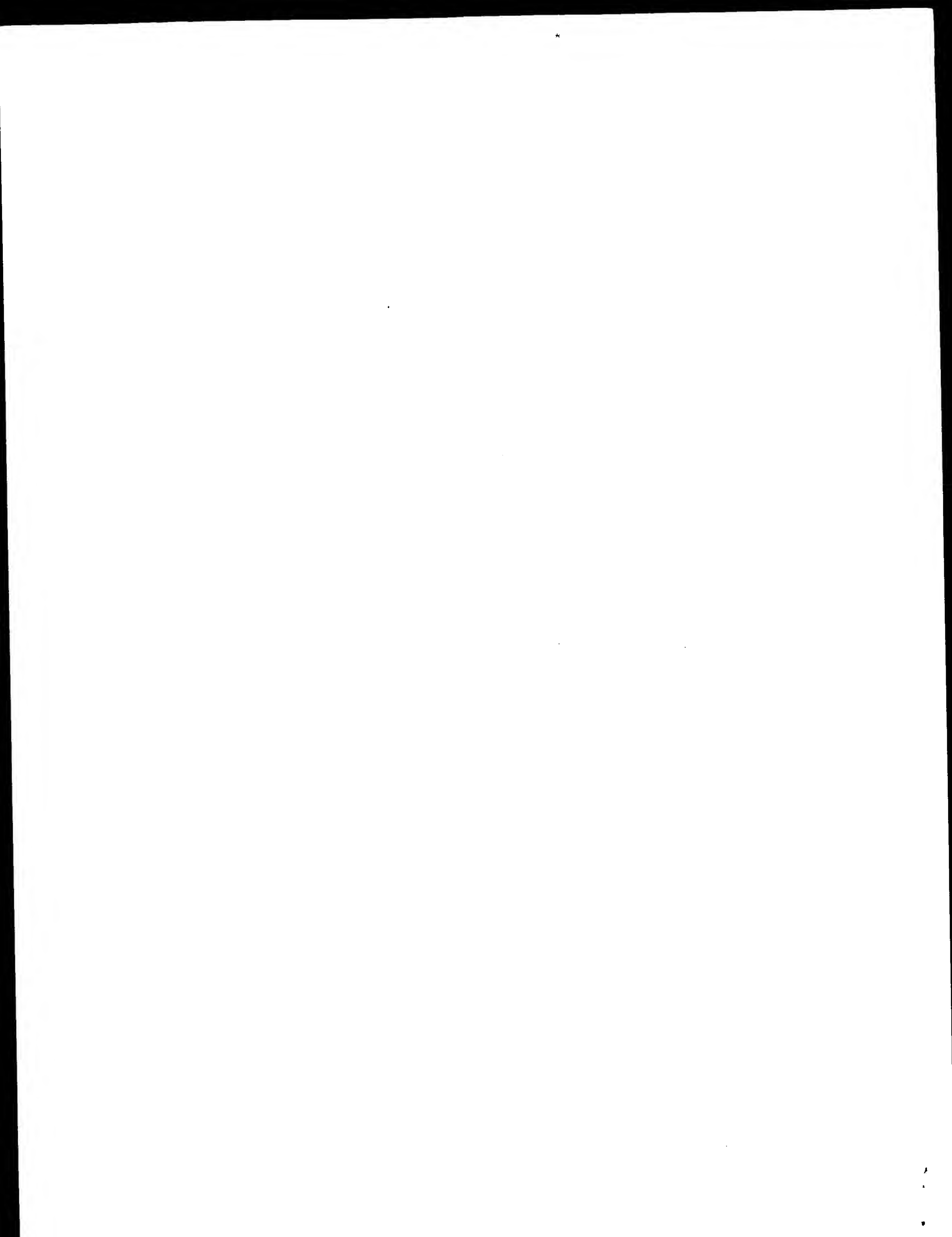
QY 210 TGGATTTGACAGCCCATCAGGGCCGGTGGTAATATTCTACAATTTGGGAAGTAAGCTGA 269
Db 349 TGGCTTTGGATAAACATATGATGTCATGTAACAATAGTTATCCCCTACCTGACAATCAGC 408
QY 270 TTGGAACAGGAAATATCAGAAGATGAAGAACTGAAGTACAAGAGCAGATGAATCTCC 329
Db 409 TACTTACCCAACTATTACGGAACCTTGAAGAAAGTCTGTAATGTTGATATGATTTTAC 468
QY 330 TAAATTCAGATGGGAATGCCTCAGGGTAGCTAGCATGGAAAAACAAGCAATTTACATA 389
Db 469 AATCGTCAACTTGGATGGCTACAACTATCCTTCAAGATTAAAGATAGTATTTCATC 528
QY 390 GAGTTTAAATGGATCTCCAGATCAGAAACTGAAAGAGTTGAATGACTGGCTAACAAAA 449
Db 529 AATTTCTATGATCTT--GAAACATTAAAGAAATCAATATGAAGCATTAGCTTATAAAAT 586
QY 450 CAGAAGAAAGAAAGAAAGAAATGGAGGAAGAGCCTCTTGGGA 490
Db 587 CAATGAGCATTATGTCAAAATTTTCATTAAATTTGGCTCAGGCA 627

RESULT 13
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc_feature
; LOCATION: (1)..(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

Query Match 1.8%; Score 36.2; DB 4; Length 289;
Best Local Similarity 6.1%; Pred. No. 0.3;
Matches 14; Conservative 99; Mismatches 118; Indels 0; Gaps 0;
QY 342 GGAATGCCTCAGGGTAGCTAGCATGGAAAAACAAGCAATTTACATAGAGTTTAAATG 401
Db 3 RGRGRARARARURURARUR 62
QY 402 ATCTCCAGATCAGAAACTGAAAGAGTTGAATGACTGGCTAACAAAAACAAGAAAGAA 461
Db 63 RNR 122
QY 462 CAAGGAAATGGAGGAAGAGCCTCTTGGACCTGATCTTGAAGACCTTAAACCGCAAGTAC 521
Db 123 RNR 182
QY 522 AACACATAAGGTGCTTCAAGAAAGATCTAGAACAAGCAAGTCAAGGTCA 572
Db 183 RNR 233

RESULT 15
 SS-09-014-969-14
 Sequence 14, Application US/09014969
 Patent No. 5965397
 GENERAL INFORMATION:
 APPLICANT: Jacobs, Kenneth
 APPLICANT: McCoy, John M.
 APPLICANT: Lavallie, Edward R.
 APPLICANT: Racie, Lisa A.
 APPLICANT: Merberg, David
 APPLICANT: Treacy, Maurice
 APPLICANT: Spaulding, Vikki
 APPLICANT: Agostino, Michael J.
 TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
 TITLE OF INVENTION: ENCODING THEM
 NUMBER OF SEQUENCES: 32
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Genetics Institute, Inc.
 STREET: 87 Cambridgepark Drive

Search completed: April 25, 2003, 17:42:33
Job time : 72.4852 secs



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OMnucleic - nucleic search, using sw model

Run on: April 25, 2003, 17:39:49 ; Search time 135.61 Seconds
(without alignments)
16376.941 Million cell updates/sec

Title: US-09-845-416-2_COPY_960_3000
Perfect score: 2041
Sequence: 1 tccttcacagcatttggaag.....acctcagcactctggaagac 2041

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 709820 seqs, 544064369 residues 1410746
Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	190.2	9.3	256	10	US-09-864-761-21956
2	180	8.8	476	10	US-09-864-761-15766
3	177	8.7	466	10	US-09-864-761-6092
4	153.4	7.5	467	10	US-09-864-761-11083
5	151	7.4	151	10	US-09-864-761-27715
6	122	6.0	122	10	US-09-864-761-32272
7	43	2.1	436	10	US-09-864-761-18355
8	42.2	2.1	423	10	US-09-864-761-18355
9	40.6	2.0	425	10	US-09-864-761-18355
10	40.4	2.0	428	10	US-09-864-761-18355
11	39.8	2.0	418	10	US-09-864-761-18355
12	39.6	1.9	1381	9	US-09-822-846-343
13	39.6	1.9	1594	9	US-10-174-590-183
14	39.6	1.9	1594	9	US-10-176-758-183
15	39.6	1.9	1594	9	US-10-175-737-183
16	39.6	1.9	1594	9	US-10-173-706-183
17	39.6	1.9	1594	9	US-10-175-738-183
18	39.6	1.9	1594	9	US-10-175-752-183
19	39.6	1.9	1594	9	US-10-176-482-183

C	20	39.6	1.9	1594	9	US-10-176-757-183	Sequence 183, App
C	21	39.6	1.9	1594	9	US-10-176-913-183	Sequence 183, App
C	22	39.6	1.9	1594	9	US-10-180-552-183	Sequence 183, App
C	23	39.6	1.9	1594	9	US-10-180-557-183	Sequence 183, App
C	24	39.6	1.9	1594	9	US-10-173-700-183	Sequence 183, App
C	25	39.6	1.9	1594	9	US-10-174-572-183	Sequence 183, App
C	26	39.6	1.9	1594	9	US-10-174-579-183	Sequence 183, App
C	27	39.6	1.9	1594	9	US-10-174-582-183	Sequence 183, App
C	28	39.6	1.9	1594	9	US-10-174-588-183	Sequence 183, App
C	29	39.6	1.9	1594	9	US-10-175-739-183	Sequence 183, App
C	30	39.6	1.9	1594	9	US-10-175-740-183	Sequence 183, App
C	31	39.6	1.9	1594	9	US-10-175-743-183	Sequence 183, App
C	32	39.6	1.9	1594	9	US-10-176-488-183	Sequence 183, App
C	33	39.6	1.9	1594	9	US-10-176-492-183	Sequence 183, App
C	34	39.6	1.9	1594	9	US-10-176-747-183	Sequence 183, App
C	35	39.6	1.9	1594	9	US-10-176-750-183	Sequence 183, App
C	36	39.6	1.9	1594	9	US-10-176-985-183	Sequence 183, App
C	37	39.6	1.9	1594	9	US-10-176-987-183	Sequence 183, App
C	38	39.6	1.9	1594	9	US-10-176-991-183	Sequence 183, App
C	39	39.6	1.9	1594	9	US-10-176-992-183	Sequence 183, App
C	40	39.6	1.9	1594	9	US-10-176-993-183	Sequence 183, App
C	41	39.6	1.9	1594	9	US-10-184-658-183	Sequence 183, App
C	42	39.6	1.9	1594	9	US-10-173-695-183	Sequence 183, App
C	43	39.6	1.9	1594	9	US-10-173-697-183	Sequence 183, App
C	44	39.6	1.9	1594	9	US-10-173-705-183	Sequence 183, App
C	45	39.6	1.9	1594	9	US-10-174-576-183	Sequence 183, App

ALIGNMENTS

RESULT 1
US-09-864-761-21956
; Sequence 21956, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/006666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/006661
; PRIOR FILING DATE: 2001-01-30

PRIOR APPLICATION NUMBER: PCT/US01/006670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 21956
LENGTH: 256
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC004468.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
OTHER INFORMATION: SWISSPROT HIT: P11532, EVALUE 2.00e-30
OTHER INFORMATION: NT HIT: g15032282, EVALUE 1.00e-103
US-09-864-761-21956

Query Match 9.3%; Score 190.2; DB 10; Length 256;
Best Local Similarity 98.5%; Pred. No. 3.3e-45;
Matches 192; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 11 CATTGGGAAGCTCCTGAAGACAAGTCAATTTGGCAGTTTCATTGATGGAGAGTGAGTAAAC 70
|||||
DB 1 CATTGGGAAGCTCCTGAAGACAAGTCAATTTGGCAGTTTCATTGATGGAGAGTGAGTAAAC 60
|||||
QY 71 CTGGACCGTTATCAAAACAGCTTTAGAGAAGATTTATCGTGGCTTCTTTCTGCTGAGGAC 130
|||||
DB 61 CTGGACCGTTATCAAAACAGCTTTAGAGAAGATTTATCGTGGCTTCTTTCTGCTGAGGAC 120
|||||
QY 131 ACATTGCAAGCACAAGGAGAGAGATTCTTAATGATGTGGAAGTGGTGAAGACCAGTTTCAT 190
|||||
DB 121 ACATTGCAAGCACAAGGAGAGAGATTCTTAATGATGTGGAAGTGGTGAAGACCAGTTTCAT 180
|||||
QY 191 ACTCATGAGGGGTAC 205
|||||
DB 181 ACTCATGAGGGTAAAC 195
|||||

RESULT 2
US-09-864-761-15766
; Sequence 15766, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27

PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 15766
LENGTH: 476
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC004468.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.41
US-09-864-761-15766

Query Match 8.8%; Score 180; DB 10; Length 476;
Best Local Similarity 100.0%; Pred. No. 4.7e-42;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 863 GTTTTAAAGCGGATCTAGAAAAGAAAAGCAATCCATGGGCAAACTGTATTCACTCAA 922
|||||
DB 294 GTTTTAAAGCGGATCTAGAAAAGAAAAGCAATCCATGGGCAAACTGTATTCACTCAA 353
|||||
QY 923 CAAGATCTTCTTTCAACACTGAAGAATAAGTCAGTGACCCAGAACGCGGATGGCTG 982
|||||
DB 354 CAAGATCTTCTTTCAACACTGAAGAATAAGTCAGTGACCCAGAACGCGGATGGCTG 413
|||||
QY 983 GATAACTTTGCCCGGTGTGGGATAATTAGTCCAAAACCTTGAAAAGAGTACAGCACAG 1042
|||||
DB 414 GATAACTTTGCCCGGTGTGGGATAATTAGTCCAAAACCTTGAAAAGAGTACAGCACAG 473
|||||

RESULT 3
US-09-864-761-6092
; Sequence 6092, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-09-27


```

; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 6092
; LENGTH: 466
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004468.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
US-09-864-761-6092

```

```

Query Match
Best Local Similarity 8.7%; Score 177; DB 10; Length 466;
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 8 CAGCATTGGAAGCTCCTGAGACAAAGTCATTGGCAGTTTCATTGATGGAGAGTGAAGTA 67
Db 290 CAGCATTGGAAGCTCCTGAGACAAAGTCATTGGCAGTTTCATTGATGGAGAGTGAAGTA 349
QY 68 AACCTGGACCGTTATCAACAGCTTTAGAAAGTATTATCGTGGCTTCTTCTGCTGAG 127
Db 350 AACCTGGACCGTTATCAACAGCTTTAGAAAGTATTATCGTGGCTTCTTCTGCTGAG 409
QY 128 GACACATTGCAAGCAGCAGAGGAGAGATTTCTAATGATGGAAGTGGTGAAGAGACCAG 184
Db 410 GACACATTGCAAGCAGCAGAGGAGAGATTTCTAATGATGGAAGTGGTGAAGAGACCAG 466

```

```

RESULT 4
US-09-864-761-11083
; Sequence 11083, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.

```

```

; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 11083
; LENGTH: 467
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004468.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.82
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.89
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.84
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.88
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.81
US-09-864-761-11083

```

```

Query Match
Best Local Similarity 7.5%; Score 153.4; DB 10; Length 467;
Matches 154; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 379 CAATTTACATAGAGTTTAAATGGATCTCCAGATCAGAACTGAAAGAGTTGAATGACTG 438
Db 313 CAGTTTACATAGAGTTTAAATGGATCTCCAGATCAGAACTGAAAGAGTTGAATGACTG 372
QY 439 GCTAACAAAAACAGAAAGAAACAAAGGAAATGGAGGAGAGCCTCTTGGACCTGATCT 498
Db 373 GCTAACAAAAACAGAAAGAAACAAAGGAAATGGAGGAGAGCCTCTTGGACCTGATCT 432
QY 499 TGAAGACCTAAACGCCAAGTACAACACATAAGG 533

```


Mon Apr 28 09:28:33 2003

Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

bb 433 TGAAGACCTAAACGCCCAAGTACAAACATAAGG 467

QY 382 TTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACTGAAAGAGTTGAATGACTGGCT 441
|||||
Db 1 TTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACTGAAAGAGTTGAATGACTGGCT 60
|||||

QY 442 AACAAAAACAGAAAGAAACAAAGGAAATGGAGGAAAGAGCCTCTTGGACCTGATCTTGA 501
|||||
Db 61 AACAAAAACAGAAAGAAACAAAGGAAATGGAGGAAAGAGCCTCTTGGACCTGATCTTGA 120
|||||
QY 502 AGACCTAAACGCCCAAGTACAAACATAAG 532
|||||
Db 121 AGACCTAAACGCCCAAGTACAAACATAAG 151
|||||

RESULT 6
US-09-864-761-32272
; Sequence 32272, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32272
; LENGTH: 122
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004468.1

US-09-864-761-27715
; Sequence 27715, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng

; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 27715
; LENGTH: 151
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004468.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.82
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.89
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.84
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.88
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.81
; OTHER INFORMATION: NT HIT: M18533.1, EVALUATE 6.00e-80
; OTHER INFORMATION: SWISSPROT HIT: P11532, EVALUATE 4.00e-08
US-09-864-761-27715

Query Match 7.4%; Score 151; DB 10; Length 151;
Best Local Similarity 100.0%; Pred. No. 6.2e-34;

OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.41
OTHER INFORMATION: EST_HUMAN HIT: BE781438.1, EVALUE 1.60e+00
OTHER INFORMATION: SWISSPROT HIT: P11532, EVALUE 1.00e-17
OTHER INFORMATION: NT HIT: X06178.1, EVALUE 9.00e-63
US-09-864-761-32272

Query Match
Best Local Similarity 100.0%; Score 122; DB 10; Length 122;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 921 AACACAGATCTTCTTCAACACTGAAGAATAAGTCAGTGACCCAGAGACGGAGCATGGC 980
Db |||||
1 AACACAGATCTTCTTCAACACTGAAGAATAAGTCAGTGACCCAGAGACGGAGCATGGC 60
QY 981 TGGATAACTTTGCCCGGTGTTGGGATAATTTAGTCCAAAAAAGTGAAGAGTACAGCAC 1040
Db |||||
61 TGGATAACTTTGCCCGGTGTTGGGATAATTTAGTCCAAAAAAGTGAAGAGTACAGCAC 120
QY 1041 AG 1042
Db ||
121 AG 122

RESULT 7
US-09-960-352-10742/c
Sequence 10742, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 10742
LENGTH: 436
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 46-LIB3058-032-Q1-K1-D10
US-09-960-352-10742

Query Match
Best Local Similarity 48.6%; Score 43; DB 10; Length 436;
Matches 118; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 274 AACAGGAAAATTATCAGAGATGAAGAAACTGAAGTACAAGACGAGATGAATCTCTTAA 333
Db |||||
431 AATAGAAAAAAGAAAGAAAAACAGACAATAAAATAAAAAATAAAAAATAAAAAATA 372
QY 334 TTCAAGATGGGAATGCCTCAGGGTAGCTAGCATGGAAAAACAAGCAATTACATAGAGT 393
Db |||||
371 TAAAGAAATAAAACAAAAATAAAATAAAATAAAATAAAATAAAATAAAATAAT 312
QY 394 TTTAATGGATCTCCAGATCAGAAACTGAAAGAGTTGAATGACTGGCTAAACAAAAACAGA 453
Db |||||
311 AAGAAGACAAAAACAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATA 252
QY 454 AGAAAGAACAGGAAATGGAGGAAGACCTCTTGGACCTGATCTTGAAGACCTAAACG 513
Db |||||
251 AAAAAAGAAAGAAAAATAAAAAAAGCATCATATAAACAAATCATTCATATATAAAAA 192
QY 514 CCA 516
Db ||
191 GCA 189

RESULT 8
US-09-864-761-18355
Sequence 18355, Application US/09864761

Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 18355
LENGTH: 423
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC010133.1
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.8
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.6
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2
US-09-864-761-18355

Query Match
Best Local Similarity 51.2%; Score 42.2; DB 10; Length 423;
Matches 148; Conservative 0; Mismatches 138; Indels 3; Gaps 2;

QY 277 AGGAAAAATTATCAGAGATGAAGAAACTGAAGTACAAGAGCA--GATGAATCTCCTAAAT 334
Db |||||
277 AGGAAAAATTATCAGAGATGAAGAAACTGAAGTACAAGAGCA--GATGAATCTCCTAAAT 334

RESULT 11

; GENERAL INFORMATION:

Query Match	2.0%	Score 39.8;	DB 10;	Length 418;
Best Local Similarity	47.1%	Pred. No. 0.32;		
Matches 122; Conservative		0; Mismatches 137;	Indels 0;	Gaps 0;

413 AAAAATAAAATAAAACAAAAAATAAAGAAAAATAAATATATAAAAAATAAG 354
 QY 340 ATGGGAATGCCTCAGGTAGCTAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAT 399
 Db 353 ACAGATAAATAAAACGAAACAAAAAATAAAAAATAGATAAAAAATAAAAAAATAAAA 294

QY 400 GGATCTCCAGAAATCAGAAACTGAAAGAGTTGAATGACTGGCTTAACAAAAACAGAGAAAG 459
Db 293 ATAAAGAAATAAAAAAGAAATAACAAAGGAAATAAAAAAGAAAAACACACACATAA 234
QY 460 AACAGGAAAAATGGAGGAAGAGCCTCTTGGACCTGATCTTGAAGACCTAAAAACGCCAAGT 519
Db 233 AAAGAGAAAAATCCAATAATCGAAACACGACGCTCGTAATAAAAAAGTCTCCAAAGCCTAGC 174
QY 520 ACAACAACATAAGCTGCTT 538
Db 173 ATCTAGACACAGAAATCCTT 155

RESULT 12

US-09-822-846-343/c
; Sequence 343, Application US/09822846
; Publication No. US20030027139A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: LaVallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Agostino, Michael J.
; APPLICANT: Steininger II, Robert J.
; APPLICANT: Bowman, Michael R.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakar
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6400
; CURRENT APPLICATION NUMBER: US/09/822,846
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,605
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 629
; SOFTWARE: PatentIn ver. 2.0
; SEQ ID NO 343
; LENGTH: 1381
; TYPE: DNA
; ORGANISM: Xenopus sp.
US-09-822-846-343

Query Match 1.9%; Score 39.6; DB 9; Length 1381;
Best Local Similarity 60.0%; Pred. No. 0.82;
Matches 66; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
QY 369 AAAAAACAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACTGAAAGAGT 428
Db 1087 AAAAAACAATTTAAATAATTTCAAAATTTTAAATTTGCTCCACCATAAGATGAATAAAGAGC 1028
QY 429 TGAATGACTGGCTAACAAAAACAGAAAGAAACAAGGAAATGGAGGAA 478
Db 1027 TTACTTAAAGGAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGGATGGA 978

RESULT 13

US-10-174-590-183/c
; Sequence 183, Application US/10174590
; Publication No. US20030008352A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C42
; CURRENT APPLICATION NUMBER: US/10/174,590
; CURRENT FILING DATE: 2002-06-18
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 183
; LENGTH: 1594
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-174-590-183

Query Match 1.9%; Score 39.6; DB 9; Length 1594;
Best Local Similarity 60.0%; Pred. No. 0.91;
Matches 66; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
QY 369 AAAAAACAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACTGAAAGAGT 428
Db 1281 AAAAAACAATTTAAATAATTTCAAAATTTTAAATTTGCTCCACCATAAGATGAATAAAGAGC 1222
QY 429 TGAATGACTGGCTAACAAAAACAGAAAGAAACAAGGAAATGGAGGAA 478
Db 1221 TTACTTAAAGGAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGGATGGA 1172

RESULT 14

US-10-176-758-183/c
; Sequence 183, Application US/10176758
; Publication No. US20030008353A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Chen, Jian
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430R1C104
; CURRENT APPLICATION NUMBER: US/10/176,758
; CURRENT FILING DATE: 2002-06-21
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 183
; LENGTH: 1594
; TYPE: DNA
; ORGANISM: Homo Sapien
US-10-176-758-183

Query Match 1.9%; Score 39.6; DB 9; Length 1594;
Best Local Similarity 60.0%; Pred. No. 0.91;
Matches 66; Conservative 0; Mismatches 44; Indels 0; Gaps 0;
QY 369 AAAAAACAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACTGAAAGAGT 428
Db 1281 AAAAAACAATTTAAATAATTTCAAAATTTTAAATTTGCTCCACCATAAGATGAATAAAGAGC 1222
QY 429 TGAATGACTGGCTAACAAAAACAGAAAGAAACAAGGAAATGGAGGAA 478
Db 1221 TTACTTAAAGGAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGGATGGA 1172

RESULT 15
 US-10-175-737-183/c
 ; Sequence 183, Application US/10175737
 ; Publication No. US20030013153A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Baker, Kevin P.
 ; APPLICANT: Chen, Jian
 ; APPLICANT: Desnoyers, Luc
 ; APPLICANT: Goddard, Audrey
 ; APPLICANT: Godowski, Paul J.
 ; APPLICANT: Gurney, Austin L.
 ; APPLICANT: Pan, James
 ; APPLICANT: Smith, Victoria
 ; APPLICANT: Watanabe, Colin K.
 ; APPLICANT: Wood, William I.
 ; APPLICANT: Zhang, Zemin
 ; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
 ; FILE REFERENCE: P3430R1C50
 ; CURRENT APPLICATION NUMBER: US/10/175,737
 ; CURRENT FILING DATE: 2002-06-19
 ; Prior Application removed - See File Wrapper or Palm
 ; NUMBER OF SEQ ID NOS: 612
 ; SEQ ID NO 183
 ; LENGTH: 1594
 ; TYPE: DNA
 ; ORGANISM: Homo Sapien
 US-10-175-737-183

Query Match 1.9%; Score 39.6; DB 9; Length 1594;
 Best Local Similarity 60.0%; Pred. No. 0.91;
 Matches 66; Conservative 0; Mismatches 44; Indels 0; Gaps 0;

Qy	369	AAAAACAAAGCAATTACATAGAGTTTAAATGGATCTCCAGATCAGAAACTGAAAGAGT	428
Db	1281	AAAAACAATTTAAATATTTCAAAATTTTAAATTTGCTCCACCATAGATGAATAAGAGC	1222
Qy	429	TGAATGACTGGCTAACAAACAGAAAGAAAGCAAGGAAATGGAGGAA	478
Db	1221	TTACTTAAGGAAAGAAAGAAACAAACAAACAAACAAACCGGATGGA	1172

Search completed: April 25, 2003, 23:58:03
 Job time : 140.776 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 07:25:57 ; Search time 284.596 Seconds
(without alignments)
16150.348 Million cell updates/sec

Title: US-09-845-416-6_COPY_1020_3060
Perfect score: 2041
Sequence: 1 tgaagtaaacctggaccgtt.....cttagctgacctgaataat 2041

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4313688

Minimum DB seq length: 0
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2041	100.0	3999	24	AAD37234
2	2041	100.0	4966	24	AAD37256
3	2041	100.0	4990	24	AAD37262
4	1749	85.7	3858	24	AAD37237
5	1749	85.7	4825	24	AAD37257
6	1749	85.7	4848	24	AAD37263
7	1702.8	83.4	4182	24	AAD37230
8	1242	60.9	2169	24	AAD37232
9	1242	60.9	3531	24	AAD37238
					Human dystrophin m
					Adeno-associated v
					Adeno-associated v
					Human dystrophin m
					Adeno-associated v
					Adeno-associated v
					Human dystrophin m
					Human dystrophin r
					Human dystrophin m

10	1242	60.9	4498	24	AAD37258	Adeno-associated v
11	1053	51.6	3510	24	AAD37240	Human dystrophin m
12	1053	51.6	4476	24	AAD37259	Adeno-associated v
13	895	43.9	1821	24	AAD37241	Human dystrophin r
14	658.4	32.3	4402	21	AAZ48568	A rod shortened dy
15	658.4	32.3	4414	24	AAD37260	Adeno-associated v
16	658	32.2	1667	24	AAD37231	Human dystrophin N
17	657	32.2	1667	24	AAD37235	Human dystrophin N
18	646.4	31.7	3446	24	AAD37242	Human dystrophin m
19	598.8	29.3	4075	21	AAZ48569	A rod shortened dy
20	594.2	29.1	4402	21	AAZ48567	A rod shortened dy
21	521.6	25.6	3275	10	AAN97129	Partial sequence o
22	508	24.9	1434	24	AAD37243	Human dystrophin r
23	387	19.0	387	24	ABK81990	Human dystrophin s
24	348	17.1	348	24	ABK81989	Human dystrophin s
25	330	16.2	333	24	ABK81967	Human dystrophin s
26	330	16.2	1340	24	AAD37239	Human dystrophin N
27	327	16.0	327	24	ABK81966	Human dystrophin s
28	324	15.9	324	24	ABK81991	Human dystrophin s
29	289.2	14.2	3747	21	AAZ48566	A rod shortened dy
30	195.2	9.6	1500	22	AAF84672	Nucleotide sequenc
31	184.6	9.0	3163	21	AAZ48571	A rod shortened dy
32	179	8.8	215	24	ABK81992	Human dystrophin h
33	153.4	7.5	467	22	ABA58159	Human foetal liver
34	153.4	7.5	467	22	AAK06239	Human bone expres
35	153.4	7.5	467	22	AAK31898	Human bone marrow
36	153.4	7.5	467	22	AAI37766	Probe #6452 used t
37	153.4	7.5	467	24	ABS06660	Human genome-deriv
38	151	7.4	151	22	ABA70760	Human foetal liver
39	151	7.4	151	22	AAK19013	Human brain expres
40	151	7.4	151	22	AAK44963	Human bone marrow
41	151	7.4	151	22	AAI50933	Probe #19619 used
42	151	7.4	151	24	ABS19208	Human genome-deriv
43	142.4	7.0	811	11	AAQ03996	Sequence complemen
44	140.8	6.9	147	24	AAD37236	Human dystrophin h
45	140.2	6.9	256	22	ABA69736	Human foetal liver

ALIGNMENTS

RESULT 1
AAD37234
ID AAD37234 standard; DNA; 3999 BP.
XX
AC AAD37234;
XX
DT 21-AUG-2002 (first entry)
XX
DE Human dystrophin minigene delta3990.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Homo sapiens.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
XX
PI Xiao X;
XX
DR WPI; 2002-049342/06.
XX
PT New dystrophin minigene for treating Duchenne or Becker muscular dystrophy comprises an N-terminal domain or modified N-terminal domain,

rod repeats, H1 and H4 domains and a cysteine rich domain of a dystrophin gene -

Example 1; Page 46-47; 71pp; English.

The present invention relates to an isolated nucleotide sequence encoding a dystrophin minigene. The minigene comprises N-terminal or modified N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control element. The dystrophin minigene in operable linkage with an expression control element, in a recombinant adeno-associated virus or retrovirus is useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is human dystrophin minigene delta3990 containing nucleotides 1-1668 (N-terminus, hinge H1 and rods R1 and R2), 8059-10227 (rods R22, R23 and R24, hinge H4 and CR domain) and 11047-11058 (dystrophin last 3 amino acids).

Sequence 3999 BP; 1223 A; 907 C; 933 G; 936 T; 0 other;

Query Match 100.0%; Score 2041; DB 24; Length 3999;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2041; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	TGAAGTAACTGGACCGTTATCAAAACAGCTTTAGAAAGAGTATTATCGTGGCTTCTTTC	60
DB	1020	TGAAGTAACTGGACCGTTATCAAAACAGCTTTAGAAAGAGTATTATCGTGGCTTCTTTC	1079
QY	61	TGCTGAGGACACATTCGAAGCACAAGGAGAGATTTCTTAATGATGTGGAACTGGTGAAGA	120
DB	1080	TGCTGAGGACACATTCGAAGCACAAGGAGAGATTTCTTAATGATGTGGAACTGGTGAAGA	1139
QY	121	CCAGTTTTCATACATGAGGGGTACATGATGGATTTGACAGCCCATCAGGGCCGGTTGG	180
DB	1140	CCAGTTTTCATACATGAGGGGTACATGATGGATTTGACAGCCCATCAGGGCCGGTTGG	1199
QY	181	TAATATTCTACAAATGGGAAGTAAGCTGATTGGAAACAGGAAAATTTATCAGAAGATGAAGA	240
DB	1200	TAATATTCTACAAATGGGAAGTAAGCTGATTGGAAACAGGAAAATTTATCAGAAGATGAAGA	1259
QY	241	AACTGAAGTACAAGAGCAGATGAATCTCTAAATTCAGATGGGAATGCCTCAGGGTAGC	300
DB	1260	AACTGAAGTACAAGAGCAGATGAATCTCTAAATTCAGATGGGAATGCCTCAGGGTAGC	1319
QY	301	TAGCATGGAAAAACAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT	360
DB	1320	TAGCATGGAAAAACAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT	1379
QY	361	GAAAGAGTTGAATGACTGGCTTAACAAAAACAGAAAGAAACAGGAAAATGGAGGAAGA	420
DB	1380	GAAAGAGTTGAATGACTGGCTTAACAAAAACAGAAAGAAACAGGAAAATGGAGGAAGA	1439
QY	421	GCCTCTTGGACCTGATCTTGAAGACCTAAAAACGCCAAGTACAACACATAAGGTGCTTCA	480
DB	1440	GCCTCTTGGACCTGATCTTGAAGACCTAAAAACGCCAAGTACAACACATAAGGTGCTTCA	1499
QY	481	AGAAGATCTAGAACAAGAACAGTCAAGGTCAATTTCTCACTCACATGGTGGGTAGT	540
DB	1500	AGAAGATCTAGAACAAGAACAGTCAAGGTCAATTTCTCACTCACATGGTGGGTAGT	1559
QY	541	TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTTGAAGAACAACACTTAAGTATTGGG	600
DB	1560	TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTTGAAGAACAACACTTAAGTATTGGG	1619
QY	601	AGATCGATGGCAACATCTGTAGATGGACAGAACCGCTGGGTTCTTTTACAAGACCA	660
DB	1620	AGATCGATGGCAACATCTGTAGATGGACAGAACCGCTGGGTTCTTTTACAAGACCA	1679
QY	661	GCCTGACCTAGCTCCTGGACTGACCACATTTGGAGCCTCTCCTACTCAGACTGTTACTCT	720
DB	1680	GCCTGACCTAGCTCCTGGACTGACCACATTTGGAGCCTCTCCTACTCAGACTGTTACTCT	1739

QY	721	GGTGACACAACTGTGGTACTAAGGAAACTGCCATCTCAAAACTAGAAAATGCCATCTTC	780
DB	1740	GGTGACACAACTGTGGTACTAAGGAAACTGCCATCTCAAAACTAGAAAATGCCATCTTC	1799
QY	781	CTTGATGTTGGAGGTACCTACTCATAGATTACTGCAACAGTTCCCTCCCTGGACCTGGAAA	840
DB	1800	CTTGATGTTGGAGGTACCTACTCATAGATTACTGCAACAGTTCCCTCCCTGGACCTGGAAA	1859
QY	841	GTTTCTTGCCTGGCTTACAGAAAGCTGAAACAACAGTCCCAATGTCTTACAGGATGCTACCCG	900
DB	1860	GTTTCTTGCCTGGCTTACAGAAAGCTGAAACAACAGTCCCAATGTCTTACAGGATGCTACCCG	1919
QY	901	TAAGGAAAGGCTCCTAGAGACTCCAAAGGAGTAAAGAGCTGATGAAACAATGGCAAGA	960
DB	1920	TAAGGAAAGGCTCCTAGAGACTCCAAAGGAGTAAAGAGCTGATGAAACAATGGCAAGA	1979
QY	961	CCTCCAAGGTGAAATTTGAAGCTCACACAGATGTTTATCACAACTGGATGAAACACAGCCA	1020
DB	1980	CCTCCAAGGTGAAATTTGAAGCTCACACAGATGTTTATCACAACTGGATGAAACACAGCCA	2039
QY	1021	AAAAATCCTGAGATCCCTGGAAAGTTCCGATGATGCAGTCTCTTACAAAGACGTTTGA	1080
DB	2040	AAAAATCCTGAGATCCCTGGAAAGTTCCGATGATGCAGTCTCTTACAAAGACGTTTGA	2099
QY	1081	TAACATGAACCTCAAGTGGAGTGAACCTTCGAAAAAGTCTCTCAACATTTAGTCCCATTT	1140
DB	2100	TAACATGAACCTCAAGTGGAGTGAACCTTCGAAAAAGTCTCTCAACATTTAGTCCCATTT	2159
QY	1141	GGAAGCCAGTTCTGACCAAGTGGAAAGCTGTGCACCTTCTCTGCAGGAACTTCTGGTGTG	1200
DB	2160	GGAAGCCAGTTCTGACCAAGTGGAAAGCTGTGCACCTTCTCTGCAGGAACTTCTGGTGTG	2219
QY	1201	GCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGACCTATTTGGAGGCGACTTTCCAGC	1260
DB	2220	GCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGACCTATTTGGAGGCGACTTTCCAGC	2279
QY	1261	AGTTCAGAAAGCAGAACGATGTACATAGGGCTTCAAGAGGGAAATTTGAAACATAAGAAC	1320
DB	2280	AGTTCAGAAAGCAGAACGATGTACATAGGGCTTCAAGAGGGAAATTTGAAACATAAGAAC	2339
QY	1321	TGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGACGACCTTTGGAAGG	1380
DB	2340	TGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGACGACCTTTGGAAGG	2399
QY	1381	ACTAGAGAAACTCTACCAGAGCCCAAGAGAGCTGCCTCCTGAGGAGAGAGAGAGAGAGT	1440
DB	2400	ACTAGAGAAACTCTACCAGAGCCCAAGAGAGCTGCCTCCTGAGGAGAGAGAGAGAGAGT	2459
QY	1441	CACCTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTTGAACCT	1500
DB	2460	CACCTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTTGAACCT	2519
QY	1501	GCACCTCCGCTGACTGGCAGAGAAAAATAGATGAGACCTTGAAGAGACTCCAGGAACTTCA	1560
DB	2520	GCACCTCCGCTGACTGGCAGAGAAAAATAGATGAGACCTTGAAGAGACTCCAGGAACTTCA	2579
QY	1561	AGAGCCACGGATGAGCTGGACCTCAAGCTGCGCAAGCTGAGGTGATCAAGGGATCCTG	1620
DB	2580	AGAGCCACGGATGAGCTGGACCTCAAGCTGCGCAAGCTGAGGTGATCAAGGGATCCTG	2639
QY	1621	GCAGCCCTGGGCGATCTCCTCAATTTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGC	1680
DB	2640	GCAGCCCTGGGCGATCTCCTCAATTTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGC	2699
QY	1681	ACTTCGAGAGAAATTTGGCCTCTGAAAGAGAGACGTGAGCCACGTCAATGACCTTGCTCG	1740
DB	2700	ACTTCGAGAGAAATTTGGCCTCTGAAAGAGAGACGTGAGCCACGTCAATGACCTTGCTCG	2759
QY	1741	CCAGCTTACCCTTTGGGCTTTCAGCTCTCACCGTATACCTCAGCACCTCTGGAAGACCT	1800
DB	2760	CCAGCTTACCCTTTGGGCTTTCAGCTCTCACCGTATACCTCAGCACCTCTGGAAGACCT	2819
QY	1801	GAACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGA	1860

Db 2820 GAACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGACCGAGTCAGGCAGCTGCATGA 2879
QY 1861 AGCCACACAGGACTTTGGTCCAGCATCTCAGCACTTCTTTCCACGCTCTGTCCAGGGTCC 1920
Db 2880 AGCCACACAGGACTTTGGTCCAGCATCTCAGCACTTCTTTCCACGCTCTGTCCAGGGTCC 2939
QY 1921 CTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCTACTATATCAACCACGAGACTCAAAC 1980
Db 2940 CTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCTACTATATCAACCACGAGACTCAAAC 2999
QY 1981 AACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAA 2040
Db 3000 AACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAA 3059
QY 2041 T 2041
Db 3060 T 3060

RESULT 2
AAD37256
ID AAD37256 standard; DNA; 4966 BP.
XX AC AAD37256;
XX 21-AUG-2002 (first entry)
DE Adeno-associated virus vector plasmid, AAV-MCK-delta3990.
XX Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX Chimeric - Homo sapiens.
OS Chimeric - Unidentified.
XX WO200183695-A2.
PN 08-NOV-2001.
XX 27-APR-2001; 2001WO-US13677.
XX 28-APR-2000; 2000US-200777P.
XX (XIAO/) XIAO X.
PA Xiao x;
XX WPI; 2002-049342/06.

DR New dystrophin minigene for treating Duchenne or Becker muscular
XX dystrophy comprises an N-terminal domain or modified N-terminal domain,
XX rod repeats, H1 and H4 domains and a cysteine rich domain of a
XX dystrophin gene -
PS Example 1; Page 59-60; 71pp; English.

XX The present invention relates to an isolated nucleotide sequence encoding
XX a dystrophin minigene. The minigene comprises N-terminal or modified
XX N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
XX domains and cysteine-rich domains of dystrophin or utrophin genes. The
XX invention also relates to a recombinant adeno-associated virus (AAV)
XX comprising dystrophin minigene operably linked to an expression control
XX element. The dystrophin minigene in operable linkage with an expression
XX control element, in a recombinant adeno-associated virus or retrovirus is
XX useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
XX dystrophy (BMD) in a mammalian subject. The present sequence is AAV
XX vector plasmid construct containing human dystrophin minigenes, a muscle
XX creatine kinase (MCK) promoter and a small polyA signal sequence.
XX Sequence 4966 BP; 1403 A; 1216 C; 1232 G; 1115 T; 0 other;

Query Match 100.0%; Score 2041; DB 24; Length 4966;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 2041; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAGTATTATCGTGGCTTCTTTC 60
Db 1777 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAGTATTATCGTGGCTTCTTTC 1836
QY 61 TGCTGAGGACACATTTGCAAGCACAAGGAGAGATTTCTTAATGATGTGGAAGTGGTGAAGA 120
Db 1837 TGCTGAGGACACATTTGCAAGCACAAGGAGAGATTTCTTAATGATGTGGAAGTGGTGAAGA 1896
QY 121 CCAGTTTCATATCTCATGAGGGGTACATGATGGATTTGACAGCCCCATCAGGGCCGGGTGG 180
Db 1897 CCAGTTTCATATCTCATGAGGGGTACATGATGGATTTGACAGCCCCATCAGGGCCGGGTGG 1956
QY 181 TAATATTCTACAATTTGGGAAGTAAGCTGATTGGAACAGGAAATTTATCAGAAGATGAAGA 240
Db 1957 TAATATTCTACAATTTGGGAAGTAAGCTGATTGGAACAGGAAATTTATCAGAAGATGAAGA 2016
QY 241 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTTCAAGATGGGAATGCCCTCAGGTAGC 300
Db 2017 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTTCAAGATGGGAATGCCCTCAGGTAGC 2076
QY 301 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT 360
Db 2077 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT 2136
QY 361 GAAAGAGTTGAATGACTGGCTAAACAAAAACAGAAGAAACAAAGGAAAAATGGAGGAAGA 420
Db 2137 GAAAGAGTTGAATGACTGGCTAAACAAAAACAGAAGAAACAAAGGAAAAATGGAGGAAGA 2196
QY 421 GCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCAAGTACAACACATAAGGTGCTTCA 480
Db 2197 GCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCAAGTACAACACATAAGGTGCTTCA 2256
QY 481 AGAAGATCTAGAACAAAGCAAGTCAAGGTCAATTTCTCTCACTCACATGCTGGTAGT 540
Db 2257 AGAAGATCTAGAACAAAGCAAGTCAAGGTCAATTTCTCTCACTCACATGCTGGTAGT 2316
QY 541 TGATGAATCTAGTGGAGATCAGCAACTGCTGCTTTGGAGACAACTTAAGGTATTGGG 600
Db 2317 TGATGAATCTAGTGGAGATCAGCAACTGCTGCTTTGGAGACAACTTAAGGTATTGGG 2376
QY 601 AGATCGATGGGCAAAACATCTGTAGATGGACAGAGACCCGCTGGTCTTTTACAAGACCA 660
Db 2377 AGATCGATGGGCAAAACATCTGTAGATGGACAGAGACCCGCTGGTCTTTTACAAGACCA 2436
QY 661 GCCTGACCTAGCTCCTGGACTGACCACCTATTGGAGCTCTCTCTACTCAGACTGTTACTCT 720
Db 2437 GCCTGACCTAGCTCCTGGACTGACCACCTATTGGAGCTCTCTCTACTCAGACTGTTACTCT 2496
QY 721 GGTGACACAACCTGTGGTTACTAAGGAAACTGCCATCTCCAAACTAGAAATGCCATCTTC 780
Db 2497 GGTGACACAACCTGTGGTTACTAAGGAAACTGCCATCTCCAAACTAGAAATGCCATCTTC 2556
QY 781 CTTGATGTTGGAGGTACCTACTCATAGATTACTGCAACAGTTCCCCCTGGACCTGGAAAA 840
Db 2557 CTTGATGTTGGAGGTACCTACTCATAGATTACTGCAACAGTTCCCCCTGGACCTGGAAAA 2616
QY 841 GTTCTTGCCTGGCTTACAGAAGCTGAAACAACTGCCAATGTCTTACAGGATGCTACCCG 900
Db 2617 GTTCTTGCCTGGCTTACAGAAGCTGAAACAACTGCCAATGTCTTACAGGATGCTACCCG 2676
QY 901 TAAGGAAAGGCTCCTAGAGACTCCAAGGGAGTAAAAAGAGCTGATGAAACAATGGCAAGA 960
Db 2677 TAAGGAAAGGCTCCTAGAGACTCCAAGGGAGTAAAAAGAGCTGATGAAACAATGGCAAGA 2736
QY 961 CCTCCAAGGTGAAATTTGAAGCTTCACACAGATGTTTATCACAACCTGGATGAAACAGCCA 1020
Db 2737 CCTCCAAGGTGAAATTTGAAGCTTCACACAGATGTTTATCACAACCTGGATGAAACAGCCA 2796
QY 1021 AAAAATCCTGAGATCCCTGGGAAGGTTCCGATGATGCAGTCTCTTACAAAGACGTTTGA 1080

||||| 2797 AAAAACTCTGAGATCCCTGGAAGGTTCCGATGATGAGTCCCTGTTACAAAGACGTTTGA 2856
1081 TAACATGAACCTTCAAGTGGAGTGAACCTTCGGAAGAGTCTCTCAACATTAGGTCCTTTC 1140
2857 TAACATGAACCTTCAAGTGGAGTGAACCTTCGGAAGAGTCTCTCAACATTAGGTCCTTTC 2916
1141 GGAAGCCAGTCTGACAGAGTGAAGCGTCTGCACCTTCTCTGCAGGAACCTTCTGGTG 1200
2917 GGAAGCCAGTCTGACAGAGTGAAGCGTCTGCACCTTCTCTGCAGGAACCTTCTGGTG 2976
1201 GCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGACCTTATTTGGAGCGGACTTCCAGC 1260
2977 GCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGACCTTATTTGGAGCGGACTTCCAGC 3036
1261 AGTTCAGAAGCAGAACGATGTACATAGGGCCTTCAAGAGGGGAAATGAAGAACTTAAAGAAC 1320
3037 AGTTCAGAAGCAGAACGATGTACATAGGGCCTTCAAGAGGGGAAATGAAGAACTTAAAGAAC 3096
1321 TGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTTGAAGG 1380
3097 TGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTTGAAGG 3156
1381 ACTAGAGAACTCTACCAGGAGCCCGAGAGAGTGCCTCCTGAGGAGAGAGCCCGAGAAATGT 1440
3157 ACTAGAGAACTCTACCAGGAGCCCGAGAGAGTGCCTCCTGAGGAGAGAGCCCGAGAAATGT 3216
1441 CACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTTGAACCT 1500
3217 CACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTTGAACCT 3276
1501 GCACCTCCGCTGAGTGGCAGAGAGAAATAGATGAGACCCCTTGAAGACTCCAGGAACTTCA 1560
3277 GCACCTCCGCTGAGTGGCAGAGAGAAATAGATGAGACCCCTTGAAGACTCCAGGAACTTCA 3336
1561 AGAGGCCACGGATGAGCTGGACCTCAAGCTGGCCCAAGCTGAGGTGATCAAGGGATCCTG 1620
3337 AGAGGCCACGGATGAGCTGGACCTCAAGCTGGCCCAAGCTGAGGTGATCAAGGGATCCTG 3396
1621 GCAGCCCGTGGCGATCTCCTCATTTGACTCTCTCAAGATCACCTCGAGAAAGTCAAGGC 1680
3397 GCAGCCCGTGGCGATCTCCTCATTTGACTCTCTCAAGATCACCTCGAGAAAGTCAAGGC 3456
1681 ACTTCGAGGAGAAATTCGCGCTCTGAAAGAGAACCTGAGCCACGTCAATGACCTTGCTCG 1740
3457 ACTTCGAGGAGAAATTCGCGCTCTGAAAGAGAACCTGAGCCACGTCAATGACCTTGCTCG 3516
1741 CCAGCTTACCACCTTTGGGCAATTCAGCTCTCACCGTATACCTCAGCACTCTGGAAGACCT 1800
3517 CCAGCTTACCACCTTTGGGCAATTCAGCTCTCACCGTATACCTCAGCACTCTGGAAGACCT 3576
1801 GAACACCCAGATGGAAGCTTCTGCAAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGA 1860
3577 GAACACCCAGATGGAAGCTTCTGCAAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGA 3636
1861 AGCCACACAGGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCAGCTCTGTCCAGGGTCC 1920
3637 AGCCACACAGGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCAGCTCTGTCCAGGGTCC 3696
1921 CTGGGAGAGAGCCATCTCGCCCAACAAAGTGCCTACTATATCAACACAGAGACTCAAAAC 1980
3697 CTGGGAGAGAGCCATCTCGCCCAACAAAGTGCCTACTATATCAACACAGAGACTCAAAAC 3756
1981 AACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAA 2040
3757 AACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAA 3816
2041 T 2041
3817 T 3817

AAD37262
ID AAD37262 standard; DNA; 4990 BP.
XX
AC AAD37262;
XX
DT 21-AUG-2002 (first entry)
XX
DE Adeno-associated virus (AAV) vector plasmid, AAV-CMV-delta3990.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Cytomegalovirus.
OS Chimeric - Unidentified.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
XX
PI Xiao X;
XX
DR WPI; 2002-049342/06.
XX
PT New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -
XX
PS Example 1; Page 67-68; 71pp; English.
XX
CC The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is AAV
CC vector plasmid construct containing human dystrophin minigenes, a
CC cytomagalovirus (CMV) promoter and a small polyA signal sequence.
XX
SQ Sequence 4990 BP; 1439 A; 1185 C; 1208 G; 1158 T; 0 other;
Query Match 100.0%; Score 2041; DB 24; Length 4990;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2041; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAGAGTATTATCGTGGCTTCTTTC 60
Db 1801 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAGAGTATTATCGTGGCTTCTTTC 1860
QY 61 TGCTGAGGACACATTGCAAGCACACAGGAGAGAGATTCTAATGATGTGGAAGTGGTGAAGA 120
Db 1861 TGCTGAGGACACATTGCAAGCACACAGGAGAGAGATTCTAATGATGTGGAAGTGGTGAAGA 1920
QY 121 CCAGTTTCATCTCATGAGGGGTACATGATGAGATTGACAGCCCATCAGGCCGGTGG 180
Db 1921 CCAGTTTCATCTCATGAGGGGTACATGATGAGATTGACAGCCCATCAGGCCGGTGG 1980
QY 181 TAATATCTACAATTGGGAAGTAAGCTGATTGGAACACAGGAAAAATTATCAGAAGATGAAGA 240
Db 1981 TAATATCTACAATTGGGAAGTAAGCTGATTGGAACACAGGAAAAATTATCAGAAGATGAAGA 2040

QY 241 AACTGAAGTACAAGACAGAGATGAATCTCCTAAATTCAGATGGGAATGCCTCAGGGTAGC 300
Db 2041 AACTGAAGTACAAGACAGAGATGAATCTCCTAAATTCAGATGGGAATGCCTCAGGGTAGC 2100
QY 301 TAGCATGGAAAAACAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAAACT 360
Db 2101 TAGCATGGAAAAACAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAAACT 2160
QY 361 GAAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGAAAGAAACAAGGAAATGGAGGAAGA 420
Db 2161 GAAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGAAAGAAACAAGGAAATGGAGGAAGA 2220
QY 421 GCCTCTTGGACCTGATCTTGAAGACCTAAAAACGCCAAGTACAACAACATAAGGTGCTTCA 480
Db 2221 GCCTCTTGGACCTGATCTTGAAGACCTAAAAACGCCAAGTACAACAACATAAGGTGCTTCA 2280
QY 481 AGAAGATCTAGAACAAAGAACAAAGTCAAGGTCAATTTCTCTCACTCACATGGTGGTGTAGT 540
Db 2281 AGAAGATCTAGAACAAAGAACAAAGTCAAGGTCAATTTCTCTCACTCACATGGTGGTGTAGT 2340
QY 541 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGG 600
Db 2341 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGG 2400
QY 601 AGATCGATGGGCAAAACATCTGTAGATGGACAGAAAGACCGCTGGGTTCTTTTACAAGACCA 660
Db 2401 AGATCGATGGGCAAAACATCTGTAGATGGACAGAAAGACCGCTGGGTTCTTTTACAAGACCA 2460
QY 661 GCCTGACCTAGCTCCTGGACTGACCACTATTGGAGCCTCTCCTACTCAGACTGTTACTCT 720
Db 2461 GCCTGACCTAGCTCCTGGACTGACCACTATTGGAGCCTCTCCTACTCAGACTGTTACTCT 2520
QY 721 GGTGACACAACCTGTGGTTACTAAGGAAACTGCCATCTCCAAACTAGAAATGCCATCTTC 780
Db 2521 GGTGACACAACCTGTGGTTACTAAGGAAACTGCCATCTCCAAACTAGAAATGCCATCTTC 2580
QY 781 CTTGATGTTGGAGGTACCTACTCATAGATTACTGCAACAGTTCGCCCTGGACCTGGAAAA 840
Db 2581 CTTGATGTTGGAGGTACCTACTCATAGATTACTGCAACAGTTCGCCCTGGACCTGGAAAA 2640
QY 841 GTTCTTGCCTGGCTTACAGAAAGCTGAAACAACTGCCAATGTCCTACAGGATGCTACCCG 900
Db 2641 GTTCTTGCCTGGCTTACAGAAAGCTGAAACAACTGCCAATGTCCTACAGGATGCTACCCG 2700
QY 901 TAAGGAAAGGCTCCTTAGAAGACTCCAAGGGAGTAAAAGAGCTGATGAAACAATGGCAAGA 960
Db 2701 TAAGGAAAGGCTCCTTAGAAGACTCCAAGGGAGTAAAAGAGCTGATGAAACAATGGCAAGA 2760
QY 961 CCTCCAAGGTGAAATGAAGCTCACACAGATGTTTATACAACTGGATGAAACAGCCA 1020
Db 2761 CCTCCAAGGTGAAATGAAGCTCACACAGATGTTTATACAACTGGATGAAACAGCCA 2820
QY 1021 AAAAATCCTGAGATCCCTGGAAGGTTCCGATGATGCAGTCCCTGTTACAAAGACGTTTGGG 1080
Db 2821 AAAAATCCTGAGATCCCTGGAAGGTTCCGATGATGCAGTCCCTGTTACAAAGACGTTTGGG 2880
QY 1081 TAACATGAACCTTCAAGTGGAGTGAACCTCGGAAAAAGTCTCTCAACATTAGGTCCCAATT 1140
Db 2881 TAACATGAACCTTCAAGTGGAGTGAACCTCGGAAAAAGTCTCTCAACATTAGGTCCCAATT 2940
QY 1141 GGAAGCCAGTTCTGACCACTGGAAGGCTGACCTTGCACCTTCTCTGCAGGAACCTTCTGGTGTG 1200
Db 2941 GGAAGCCAGTTCTGACCACTGGAAGGCTGACCTTGCACCTTCTCTGCAGGAACCTTCTGGTGTG 3000
QY 1201 GCTACAGCTGAAAGATGATGAATTAAGCCGCGCAGGCACCTATTGGAGGCGACTTTCACAGC 1260
Db 3001 GCTACAGCTGAAAGATGATGAATTAAGCCGCGCAGGCACCTATTGGAGGCGACTTTCACAGC 3060
QY 1261 AGTTCAGAAGCAGAACGATGTACATAGGGCCTTCAAGAGGGGAATTGAAAACTAAACAACC 1320
Db 3061 AGTTCAGAAGCAGAACGATGTACATAGGGCCTTCAAGAGGGGAATTGAAAACTAAACAACC 3120
QY 1321 TGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAAGG 1380

Db 3121 TGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGAAGG 3180
QY 1381 ACTAGAGAAACTCTACCAGGAGCCAGAGAGCTGCTCCTGAGGAGAGAGCCAGAAATGT 1440
Db 3181 ACTAGAGAAACTCTACCAGGAGCCAGAGAGCTGCTCCTGAGGAGAGAGCCAGAAATGT 3240
QY 1441 CACTCGGCTTCTACGAAAGCAGGCTGAGGAGTCAATACTGAGTGGGAAAAATTTGAACCT 1500
Db 3241 CACTCGGCTTCTACGAAAGCAGGCTGAGGAGTCAATACTGAGTGGGAAAAATTTGAACCT 3300
QY 1501 GCCTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGACTCCAGGAACTTCA 1560
Db 3301 GCCTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGACTCCAGGAACTTCA 3360
QY 1561 AGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATCCTG 1620
Db 3361 AGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATCCTG 3420
QY 1621 GCAGCCCGTGGCGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGC 1680
Db 3421 GCAGCCCGTGGCGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGC 3480
QY 1681 ACTTCGAGGAGAAATTTGGCCTCTGAAAGAGAACTGAGCCACGTCATGACCTTGCTCG 1740
Db 3481 ACTTCGAGGAGAAATTTGGCCTCTGAAAGAGAACTGAGCCACGTCATGACCTTGCTCG 3540
QY 1741 CCAGCTTACCACCTTTGGSCATTCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCT 1800
Db 3541 CCAGCTTACCACCTTTGGSCATTCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCT 3600
QY 1801 GAACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGAGCCGAGTCAGGAGCTGCATGA 1860
Db 3601 GAACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGAGCCGAGTCAGGAGCTGCATGA 3660
QY 1861 AGCCACACAGGGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCAGCTGTCCAGGGTCC 1920
Db 3661 AGCCACACAGGGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCAGCTGTCCAGGGTCC 3720
QY 1921 CTGGGAGAGAGCCATCTCGCCAAACAAAGTCCCTACTATATCAACACGAGACTCAAAC 1980
Db 3721 CTGGGAGAGAGCCATCTCGCCAAACAAAGTCCCTACTATATCAACACGAGACTCAAAC 3780
QY 1981 AACTTGCTGGACCATCCCAAAATGACAGAGCTCTACCAAGTCTTAGCTGACCTGAATAA 2040
Db 3781 AACTTGCTGGACCATCCCAAAATGACAGAGCTCTACCAAGTCTTAGCTGACCTGAATAA 3840
QY 2041 T 2041
Db 3841 T 3841

RESULT 4
AAD37237
ID AAD37237 standard; DNA; 3858 BP.
XX
AC AAD37237;
XX
DT 21-AUG-2002 (first entry)
XX
DE Human dystrophin minigene delta3849.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
OS Homo sapiens.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.

28-APR-2000; 2000US-200777P.
(XIAO/) XIAO X.
Xiao X;
WPI; 2002-049342/06.
New dystrophin minigene for treating Duchenne or Becker muscular dystrophy comprises an N-terminal domain or modified N-terminal domain, rod repeats, H1 and H4 domains and a cysteine rich domain of a dystrophin gene -
Example 1; Page 48-49; 71pp; English.
The present invention relates to an isolated nucleotide sequence encoding a dystrophin minigene. The minigene comprises N-terminal or modified N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control element. The dystrophin minigene in operable linkage with an expression control element, in a recombinant adeno-associated virus or retrovirus is useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is human dystrophin minigene delta3849 containing nucleotides 1-1668 (N-terminus, hinge H1 and rods R1, R2), 8059-10227 (rods R22, R23 and R24, hinge H4 and CR domain) and 11047-11058 (dystrophin last 3 amino acids).
Sequence 3858 BP; 1189 A; 866 C; 905 G; 898 T; 0 other;
Query Match 85.7%; Score 1749; DB 24; Length 3858;
Best Local Similarity 93.1%; Pred. No. 0;
Matches 1900; Conservative 0; Mismatches 0; Indels 141; Gaps 1;
1 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAAGTATTATCGTGGCTTCTTTC 60
1020 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAAGTATTATCGTGGCTTCTTTC 1079
61 TGCTGAGGACACATTGCAAGCACAAAGGAGAGATTCTTAATGATGTGGAAAGTGGTGAAGA 120
1080 TGCTGAGGACACATTGCAAGCACAAAGGAGAGATTCTTAATGATGTGGAAAGTGGTGAAGA 1139
121 CCAGTTTCATCTCATGAGGGGTACATGATGGATTGACAGCCCATCAGGGCCGGGTGG 180
1140 CCAGTTTCATCTCATGAGGGGTACATGATGGATTGACAGCCCATCAGGGCCGGGTGG 1199
181 TAATATTCTACAATTGGGAAGTAAGCTGATTGGAACAGAGAAAAATTATCAGAAGATGAAGA 240
1200 TAATATTCTACAATTGGGAAGTAAGCTGATTGGAACAGAGAAAAATTATCAGAAGATGAAGA 1259
241 AACTGAAGTACAAGACAGATGAATCTCCTAAATTCAGATGGGAATGCCTCAGGGTAGC 300
1260 AACTGAAGTACAAGACAGATGAATCTCCTAAATTCAGATGGGAATGCCTCAGGGTAGC 1319
301 TAGCATGGAAAAACAAGCAATTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT 360
1320 TAGCATGGAAAAACAAGCAATTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT 1379
361 GAAAGAGTTGAATGACTGGCTTAACAAAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAA 420
1380 GAAAGAGTTGAATGACTGGCTTAACAAAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAA 1439
421 GCCTCTTGACCTGATCTTGAAGACCTAAACGCCAAGTACAAACATTAAGGTGCTTCA 480
1440 GCCTCTTGACCTGATCTTGAAGACCTAAACGCCAAGTACAAACATTAAGGTGCTTCA 1499
481 AGAAGATCTAGAACAAGAACAAAGTCAGGGTCAATTCTCTCACTACATGGTGGTAGT 540
1500 AGAAGATCTAGAACAAGAACAAAGTCAGGGTCAATTCTCTCACTACATGGTGGTAGT 1559
541 TGATGAATCTAGTGGAGATCACGCCAACTGCTGCTTTGGAAAGAACAACTTAAGGTATTGGG 600

||||| 1560 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAAGAACAACTTAAGGTATTGGG 1619
QY 601 AGATCGATGGCAAAACATCTGTAGATGGACAGAGACCGGTGGTTCTTTTACAAGACCA 660
Db 1620 AGATCGATGGGCAAAACATCTGTAGATGGACAGAGACCGGTGGTTCTTTTACAAGAC-- 1677
QY 661 GCCTGACCTAGCTCCTGGACTGACCACCTATTGGAGCCCTCTCCTACTCAGACTGTTACTCT 720
Db 1678 ----- 1677
QY 721 GGTGACACAAACCTGTGTTACTAAGGAAACTGCCATCTCCAAACTAGAAATGCCATCTTC 780
Db 1678 ----- 1677
QY 781 CTTGATGTTGGAGGTACCTACTACTATAGATTACTGCAACAGTTTCCCTCCCTGGACCTGGAAAA 840
Db 1678 -----ACTCATAGATTACTGCAACAGTTTCCCTCCCTGGACCTGGAAAA 1718
QY 841 GTTCTTGCCTGGCTTACAGAAAGCTGAAACAACTGCCAATGTCTCTACAGGATGCTACCCG 900
Db 1719 GTTCTTGCCTGGCTTACAGAAAGCTGAAACAACTGCCAATGTCTCTACAGGATGCTACCCG 1778
QY 901 TAAGGAAAGGCTCCTAGAAAGCTCAAGGGAGTAAAGAGCTGATGAAACAAATGGCAAGA 960
Db 1779 TAAGGAAAGGCTCCTAGAAAGCTCAAGGGAGTAAAGAGCTGATGAAACAAATGGCAAGA 1838
QY 961 CCTCAAGGTGAAATTGAAGCTCACACAGATGTTTATCACAACCTGGATGAAACACAGCCA 1020
Db 1839 CCTCAAGGTGAAATTGAAGCTCACACAGATGTTTATCACAACCTGGATGAAACACAGCCA 1898
QY 1021 AAAATCCTGAGATCCCTGGAAAGTTCCGATGATGCAGTCTCTGTACAAAGACGTTTGGGA 1080
Db 1899 AAAATCCTGAGATCCCTGGAAAGTTCCGATGATGCAGTCTCTGTACAAAGACGTTTGGGA 1958
QY 1081 TAACATGAACCTCAAGTGGAGTGAACCTCGGAAAAAGTCTCTCAACATTAGTCCCATTT 1140
Db 1959 TAACATGAACCTCAAGTGGAGTGAACCTCGGAAAAAGTCTCTCAACATTAGTCCCATTT 2018
QY 1141 GGAAGCCAGTTCTGACCCAGTGGAAAGCGTCTGCACCTTTCTCTGCAGGAACTTCTGGTGTG 1200
Db 2019 GGAAGCCAGTTCTGACCCAGTGGAAAGCGTCTGCACCTTTCTCTGCAGGAACTTCTGGTGTG 2078
QY 1201 GCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGACCTATTGGAGGCGACTTTCACAGC 1260
Db 2079 GCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGACCTATTGGAGGCGACTTTCACAGC 2138
QY 1261 AGTTCAAGAGCAGAACCGATGTACATAGGGCCCTTCAAGAGGGAATTGAAACATAAAGAAC 1320
Db 2139 AGTTCAAGAGCAGAACCGATGTACATAGGGCCCTTCAAGAGGGAATTGAAACATAAAGAAC 2198
QY 1321 TGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAAGG 1380
Db 2199 TGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAAGG 2258
QY 1381 ACTAGAGAAACTCTACCAGGAGCCAGAGAGCTGCCCTCCTGAGGAGAGAGAGAGAGAGATGT 1440
Db 2259 ACTAGAGAAACTCTACCAGGAGCCAGAGAGCTGCCCTCCTGAGGAGAGAGAGAGAGAGATGT 2318
QY 1441 CACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGAAAAAATTGAACCT 1500
Db 2319 CACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGAAAAAATTGAACCT 2378
QY 1501 GCACTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGACTCCAGGAACTTCA 1560
Db 2379 GCACTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGACTCCAGGAACTTCA 2438
QY 1561 AGAGCCACGGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATCCTG 1620
Db 2439 AGAGCCACGGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATCCTG 2498
QY 1621 GCAGCCCGTGGCGGATCTCCTCATTTGACTCTCTCCTCAAGATCACCTCGAGAAAGTCAAGGC 1680
|||||

Db 2499 GCAGCCCGTGGCGGATCTCCTCATTTGACTCTCTCCAAAGATCACCTCGAGAAAGTCAAGGC 2558
QY 1681 ACTTCGAGGAGAAATTCGCCCTCTGAAGAGAACGTTGAGCCACGTCAATGACCTTGCTCG 1740
Db 2559 ACTTCGAGGAGAAATTCGCCCTCTGAAGAGAACGTTGAGCCACGTCAATGACCTTGCTCG 2618
QY 1741 CCAGCTTACCACCTTTGGGCATTACGCTCTCACCGTATACCTCAGCACTCTGGAAGACCT 1800
Db 2619 CCAGCTTACCACCTTTGGGCATTACGCTCTCACCGTATACCTCAGCACTCTGGAAGACCT 2678
QY 1801 GAACACCAGATGGAAGCTTTCGAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGA 1860
Db 2679 GAACACCAGATGGAAGCTTTCGAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGA 2738
QY 1861 AGCCACACAGGACTTTGGTCCAGCATCTCAGCACCTTTCTTTCCACGCTCTGTCCAGGGTCC 1920
Db 2739 AGCCACACAGGACTTTGGTCCAGCATCTCAGCACCTTTCTTTCCACGCTCTGTCCAGGGTCC 2798
QY 1921 CTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACACGAGACTCAAAAC 1980
Db 2799 CTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACACGAGACTCAAAAC 2858
QY 1981 AACTTGCTGGGACATCCCAAAATGACAGAGCTCTACCGACTTTTAGCTGACCTGAATAA 2040
Db 2859 AACTTGCTGGGACATCCCAAAATGACAGAGCTCTACCGACTTTTAGCTGACCTGAATAA 2918
QY 2041 T 2041
Db 2919 T 2919

RESULT 5
AAD37257
ID AAD37257 standard; DNA: 4825 BP.
XX
AC AAD37257;
XX
CT 21-AUG-2002 (first entry)
XX
DE Adeno-associated virus vector plasmid, AAV-MCK-delta3849.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
DS Chimeric - Homo sapiens.
DS Chimeric - Unidentified.
XX
PN W0200183695-A2.
XX
XX 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
PA
TI Xiao X;
TI
XR WPI; 2002-049342/06.
XR
X New dystrophin minigene for treating Duchenne or Becker muscular
X dystrophy comprises an N-terminal domain or modified N-terminal domain,
X rod repeats, H1 and H4 domains and a cysteine rich domain of a
X dystrophin gene -
X
S Example 1; Page 61-62; 71pp; English.
S
X The present invention relates to an isolated nucleotide sequence encoding
X a dystrophin minigene. The minigene comprises N-terminal or modified
X N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
X domains and cysteine-rich domains of dystrophin or utrophin genes. The

CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is AAV
CC vector plasmid construct containing human dystrophin minigenes, a muscle
CC creatine kinase (MCK) promoter and a small polyA signal sequence.
XX
SQ Sequence 4825 BP; 1369 A; 1175 C; 1204 G; 1077 T; 0 other;
Query Match 85.7%; Score 1749; DB 24; Length 4825;
Best Local Similarity 93.1%; Pred. No. 0;
Matches 1900; Conservative 0; Mismatches 0; Indels 141; Gaps 1;
QY 1 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAGTATTATCGTGGCTTCTTTC 60
Db 1777 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAGTATTATCGTGGCTTCTTTC 1836
QY 61 TGCTGAGGACACATTCGAAGCACAAAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 120
Db 1837 TGCTGAGGACACATTCGAAGCACAAAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 1896
QY 121 CCAGTTTCATCTACTCATGAGGGGTACATGATGGATTTCACAGCCCATCAGGCCGGTGG 180
Db 1897 CCAGTTTCATCTACTCATGAGGGGTACATGATGGATTTCACAGCCCATCAGGCCGGTGG 1956
QY 181 TAATATCTACAATTTGGGAAGTAAGCTGATTGGAAACAGGAAATTTATCAGAAGATGAAGA 240
Db 1957 TAATATCTACAATTTGGGAAGTAAGCTGATTGGAAACAGGAAATTTATCAGAAGATGAAGA 2016
QY 241 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGATGGGAATGCCTCAGGGTAGC 300
Db 2017 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGATGGGAATGCCTCAGGGTAGC 2076
QY 301 TAGCATGGAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAAACT 360
Db 2077 TAGCATGGAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAAACT 2136
QY 361 GAAAGAGTTGAATGACTGGCTAACAAAACAGAAAGAAACAAAGGAAATGGAGGAAGA 420
Db 2137 GAAAGAGTTGAATGACTGGCTAACAAAACAGAAAGAAACAAAGGAAATGGAGGAAGA 2196
QY 421 GCCTCTTGGACCTGATCTTGAAGACCTAAACGCCAAGTACAACAACATAAGGTGCTTCA 480
Db 2197 GCCTCTTGGACCTGATCTTGAAGACCTAAACGCCAAGTACAACAACATAAGGTGCTTCA 2256
QY 481 AGAAGATCTAGAACAAGAACAAAGTCAAGGTCAATTTCTCTCACTCACATGGTGGTAGT 540
Db 2257 AGAAGATCTAGAACAAGAACAAAGTCAAGGTCAATTTCTCTCACTCACATGGTGGTAGT 2316
QY 541 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTTGGAAAGAACAACTTAAGGTATGGG 600
Db 2317 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTTGGAAAGAACAACTTAAGGTATGGG 2376
QY 601 AGATCGATGGGCAACATCTGTAGATGGACAGAGACCGCTGGGTTCTTTTACAAGACCA 660
Db 2377 AGATCGATGGGCAACATCTGTAGATGGACAGAGACCGCTGGGTTCTTTTACAAGAC-- 2434
QY 661 GCCTGACCTAGCTCCTGGACTGACCACTATTGGAGCCCTCTCTACTCAGACTGTTACTCT 720
Db 2435 ----- 2434
QY 721 GGTGACACAACTGTGGTTACTAAGGAAACTGCCATCTCCAAACTAGAAATGCCATCTTC 780
Db 2435 ----- 2434
QY 781 CTTGATGTTGGAGGTACCTACTCATAGATTACTGCAACAGTTCCCTCGACCTGGAAAA 840
Db 2435 -----ACTCATAGATTACTGCAACAGTTCCCTCGACCTGGAAAA 2475
QY 841 GTTCTTGCCTGGCTTACAGAAAGCTGAAACAACTGCCAATGTCTACAGGATGCTACCCG 900
|||||

1b 2476 GTTCTTGGCTTACAGAAGCTGAAACAACTGCCAATGTCTCTACAGGATGCTACCCG 2535
1y 901 TAAGGAAAGGCTCCTAGAAAGACTCCAAGGGAGTAAAAGAGCTGATGAAACAAATGGCAAGA 960
1b 2536 TAAGGAAAGGCTCCTAGAAAGACTCCAAGGGAGTAAAAGAGCTGATGAAACAAATGGCAAGA 2595
1y 961 CCTCCAAGGTGAATTAAGCTCACACAGATGTTTATCACAAACCTGGATGAAACACAGCCA 1020
1b 2596 CCTCCAAGGTGAATTAAGCTCACACAGATGTTTATCACAAACCTGGATGAAACACAGCCA 2655
2y 1021 AAAAATCCTGAGATCCCTGGAAGGTTCCGATGATGCCAGTCTCTTACAAAAGACGTTTGA 1080
1b 2656 AAAAATCCTGAGATCCCTGGAAGGTTCCGATGATGCCAGTCTCTTACAAAAGACGTTTGA 2715
2y 1081 TAAACATGAACCTTCAAGTGAGTGAAGTTCGGAAAAAGTCTCTCAACATTAGTCCCATTT 1140
1b 2716 TAAACATGAACCTTCAAGTGAGTGAAGTTCGGAAAAAGTCTCTCAACATTAGTCCCATTT 2775
2y 1141 GGAAGCCAGTTCTGACCAAGTGGAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGTG 1200
1b 2776 GGAAGCCAGTTCTGACCAAGTGGAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGTG 2835
2y 1201 GCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACTTATGAGGCGGACTTCCAGC 1260
1b 2836 GCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACTTATGAGGCGGACTTCCAGC 2895
2y 1261 AGTTCAAGACGAGAACGATGTACATFAGGGCCCTTCAAGAGGGAAATGAAAACTAAAGAAC 1320
1b 2896 AGTTCAAGACGAGAACGATGTACATFAGGGCCCTTCAAGAGGGAAATGAAAACTAAAGAAC 2955
2y 1321 TGTATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGAAAG 1380
1b 2956 TGTATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGAAAG 3015
2y 1381 ACTAGAGAACTCTTACCAGGAGCCAGAGAGTGCCTCCTGAGGAGAGAGCCAGCAATGT 1440
1b 3016 ACTAGAGAACTCTTACCAGGAGCCAGAGAGTGCCTCCTGAGGAGAGAGCCAGCAATGT 3075
2y 1441 CACTCGGCTTTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATGAAACCT 1500
1b 3076 CACTCGGCTTTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATGAAACCT 3135
2y 1501 GCACCTCCGCTGACTGGCAGAGAAAAATAGATGAGACCTTGAAGAGACTCCAGGAATTCA 1560
1b 3136 GCACCTCCGCTGACTGGCAGAGAAAAATAGATGAGACCTTGAAGAGACTCCAGGAATTCA 3195
2y 1561 AGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCAGCTGAGGTGATCAAGGGATCCTG 1620
1b 3196 AGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCAGCTGAGGTGATCAAGGGATCCTG 3255
2y 1621 GCAGCCCGTGGCGGATCCTCATTTGACTCTCTCCAAAGATCACCTCGAGAAAGTCAAGGC 1680
1b 3256 GCAGCCCGTGGCGGATCCTCATTTGACTCTCTCCAAAGATCACCTCGAGAAAGTCAAGGC 3315
2y 1681 ACTTCGAGGAGAAATTCGCGCTCTGAAAGAGAACGTGAGCCACGTCAATGACCTTGCTCG 1740
1b 3316 ACTTCGAGGAGAAATTCGCGCTCTGAAAGAGAACGTGAGCCACGTCAATGACCTTGCTCG 3375
2y 1741 CCAGCTTACCACCTTTGGGCATTCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCT 1800
1b 3376 CCAGCTTACCACCTTTGGGCATTCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCT 3435
2y 1801 GAACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGCAGCTGCATGA 1860
1b 3436 GAACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGCAGCTGCATGA 3495
2y 1861 AGCCACAGGGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGTCTGTCAGGGTCC 1920
1b 3496 AGCCACAGGGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGTCTGTCAGGGTCC 3555
2y 1921 CTGGGAGAGAGCCCATCTCGCCCAACAAAGTGCCTTACTATATCAACCACGAGACTCAAAAC 1980
1b 3556 CTGGGAGAGAGCCCATCTCGCCCAACAAAGTGCCTTACTATATCAACCACGAGACTCAAAAC 3615

QY 1981 AACTTGTGGACCATCCCAAAATGACAGAGCTCTACAGCTCTTTAGCTGACCTGAATAA 2040
Db 3616 AACTTGTGGACCATCCCAAAATGACAGAGCTCTACAGCTCTTTAGCTGACCTGAATAA 3675
QY 2041 T 2041
Db 3676 T 3676
RESULT 6
AAD37263
ID AAD37263 standard; DNA; 4848 BP.
XX
AC AAD37263;
XX
DT 21-AUG-2002 (first entry)
XX
DE Adeno-associated virus (AAV) vector plasmid, AAV-CMV-delta3849.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
Becker muscular dystrophy; ds.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Cytomegalovirus.
OS Chimeric - Unidentified.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
XX
PI Xiao X;
XX
DR WPI; 2002-049342/06.
XX
PT New dystrophin minigene for treating Duchenne or Becker muscular
dystrophy comprises an N-terminal domain or modified N-terminal domain,
rod repeats, H1 and H4 domains and a cysteine rich domain of a
dystrophin gene -
PT
XX
PS Example 1; Page 68-70; 71pp; English.
XX
CC The present invention relates to an isolated nucleotide sequence encoding
a dystrophin minigene. The minigene comprises N-terminal or modified
N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
domains and cysteine-rich domains of dystrophin or utrophin genes. The
invention also relates to a recombinant adeno-associated virus (AAV)
comprising dystrophin minigene operably linked to an expression control
element. The dystrophin minigene in operable linkage with an expression
control element, in a recombinant adeno-associated virus or retrovirus is
useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
dystrophy (BMD) in a mammalian subject. The present sequence is AAV
vector plasmid construct containing human dystrophin minigenes, a
cytomegalovirus (CMV) promoter and a small polyA signal sequence.
XX
SQ Sequence 4848 BP; 1405 A; 1144 C; 1180 G; 1119 T; 0 other;
Query Match 85.7%; Score 1749; DB 24; Length 4848;
Best Local Similarity 93.1%; Pred. No. 0;
Matches 1900; Conservative 0; Mismatches 0; Indels 141; Gaps 1;
QY 1 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAGAAGTATTATCGTGGCTCTTTC 60
Db 1800 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAGAAGTATTATCGTGGCTCTTTC 1859
QY 61 TGCTGAGGACACATTCGAAGCACCAAGGAGAGATTTCTAATGATGTGGAAGTGTCGAAGA 120

Db 1860 TGCTGAGGACACATTGCAAGCACAAAGGAGAGATTCTTAATGATGTGGAAGTGGTGAAGA 1919
QY 121 CCAGTTTCATACTCATGAGGGGTACATGATGGATTTGACAGCCCATCAGGGCCGGGTGG 180
Db 1920 CCAGTTTCATACTCATGAGGGGTACATGATGGATTTGACAGCCCATCAGGGCCGGGTGG 1979
QY 181 TAATATTCTACAATTGGGAAGTAAGCTGATTGGAACAGGAAAAATTATCAGAAGATGAAGA 240
Db 1980 TAATATTCTACAATTGGGAAGTAAGCTGATTGGAACAGGAAAAATTATCAGAAGATGAAGA 2039
QY 241 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAAGATGGGAATGCCCTCAGGGTAGC 300
Db 2040 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAAGATGGGAATGCCCTCAGGGTAGC 2099
QY 301 TAGCATGGAAAAACAAGCAATTTACATAGAGTTTAAATGGATCTCAGAAATCAGAAAAT 360
Db 2100 TAGCATGGAAAAACAAGCAATTTACATAGAGTTTAAATGGATCTCAGAAATCAGAAAAT 2159
QY 361 GAAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGAAAGAAAGAAATGGAGGAAGA 420
Db 2160 GAAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGAAAGAAAGAAATGGAGGAAGA 2219
QY 421 GCCTCTTGGACCTGATCTTGAAGACCTAAAAACGCCAAGTACAACAACATAAGGTGCTTCA 480
Db 2220 GCCTCTTGGACCTGATCTTGAAGACCTAAAAACGCCAAGTACAACAACATAAGGTGCTTCA 2279
QY 481 AGAAGATCTAGAACAAGAACAAAGTCAAGGTCAATTTCTCACTCACATGCTGGTGGTAGT 540
Db 2280 AGAAGATCTAGAACAAGAACAAAGTCAAGGTCAATTTCTCACTCACATGCTGGTGGTAGT 2339
QY 541 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGG 600
Db 2340 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGG 2399
QY 601 AGATCGATGGGCAAAACATCTGTAGATGGACAGAACCCGCTGGGTTCTTTTACAAGACCA 660
Db 2400 AGATCGATGGGCAAAACATCTGTAGATGGACAGAACCCGCTGGGTTCTTTTACAAGAC -- 2457
QY 661 GCCTGACCTAGCTCCTGGACTGACCACTATTGGAGCCCTCTCCTACTCAGACTGTTACTCT 720
Db 2458 ----- 2457
QY 721 GGTGACACAACCTGTGGTTACTAAGGAAACTGCCATCTCCAAACTAGAAATGCCATCTTC 780
Db 2458 ----- 2457
QY 781 CTTGATGTTGGAGGTACCTACTCATAGATTACTGCAACAGTTCCCCCTGGACCTGGAAAA 840
Db 2458 -----ACTCATAGATTACTGCAACAGTTCCCCCTGGACCTGGAAAA 2498
QY 841 GTTTCTTGCCTGGCTTACAGAAGCTGAAACAACCTGCCAATGTCTCTACAGGATGCTACCCG 900
Db 2499 GTTTCTTGCCTGGCTTACAGAAGCTGAAACAACCTGCCAATGTCTCTACAGGATGCTACCCG 2558
QY 901 TAAGGAAAGGCTCCTAGAAAGACTCCAAGGGAGTAAAAGAGCTGATGAAACAATGGCAAGA 960
Db 2559 TAAGGAAAGGCTCCTAGAAAGACTCCAAGGGAGTAAAAGAGCTGATGAAACAATGGCAAGA 2618
QY 961 CCTCCAAGGTGAATTAAGCTCACACAGATGTTTATCACAACTGGATGAAACAACAGCCA 1020
Db 2619 CCTCCAAGGTGAATTAAGCTCACACAGATGTTTATCACAACTGGATGAAACAACAGCCA 2678
QY 1021 AAAATCCTGAGATCCCTGGAAGGTTCCGATGATGCAGTCCCTGTTTACAAGACGTTTGGG 1080
Db 2679 AAAATCCTGAGATCCCTGGAAGGTTCCGATGATGCAGTCCCTGTTTACAAGACGTTTGGG 2738
QY 1081 TAACATGAACCTCAAGTGGAGTGAACCTTCGGAAGAAAGTCTCTCAACATTAGGTCCCAATT 1140
Db 2739 TAACATGAACCTCAAGTGGAGTGAACCTTCGGAAGAAAGTCTCTCAACATTAGGTCCCAATT 2798
QY 1141 GGAAGCCAGTTCTGACCAGTGAAGCGCTTCACACCTTTCTCTGCAGGAACCTTCTGGTGTG 1200
Db -----

Db 2799 GGAAGCCAGTTTCTGACCAGTGAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGTG 2858
QY 1201 GCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTGGAGGCGACTTTCCAGC 1260
Db 2859 GCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTGGAGGCGACTTTCCAGC 2918
QY 1261 AGTTCAGAAGCAGAACGATGTACATAGGGCCCTTCAAGAGGGAATTTGAAAACTAAAGAACC 1320
Db 2919 AGTTCAGAAGCAGAACGATGTACATAGGGCCCTTCAAGAGGGAATTTGAAAACTAAAGAACC 2978
QY 1321 TGTAAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGAAGG 1380
Db 2979 TGTAAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGAAGG 3038
QY 1381 ACTAGAGAAAACTCTACAGGAGCCAGAGAGCTGCCTCCTGAGGAGAGAGCCAGAAATGT 1440
Db 3039 ACTAGAGAAAACTCTACAGGAGCCAGAGAGCTGCCTCCTGAGGAGAGAGCCAGAAATGT 3098
QY 1441 CACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGAAAAATTTGAACCT 1500
Db 3099 CACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGAAAAATTTGAACCT 3158
QY 1501 GCACTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGACTCCAGGAACCTCA 1560
Db 3159 GCACTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGACTCCAGGAACCTCA 3218
QY 1561 AGAGGCCACGGATGAGCTGGACCTCAAGCTCGGCCAAGCTGAGGATGATCAAGGGATCCTG 1620
Db 3219 AGAGGCCACGGATGAGCTGGACCTCAAGCTCGGCCAAGCTGAGGATGATCAAGGGATCCTG 3278
QY 1621 GCAGCCCGTGGGCGATCTCCTCATTTGACTCTCTCCAGATCACCTCGAGAAAGTCAAGGC 1680
Db 3279 GCAGCCCGTGGGCGATCTCCTCATTTGACTCTCTCCAGATCACCTCGAGAAAGTCAAGGC 3338
QY 1681 ACTTCGAGGAGAAAAATTCGCGCTCTGAAAGAGAACGCTGAGCCACGTCATGACCTTGCTCG 1740
Db 3339 ACTTCGAGGAGAAAAATTCGCGCTCTGAAAGAGAACGCTGAGCCACGTCATGACCTTGCTCG 3398
QY 1741 CCAGCTTACCACCTTTGGGCATTCAGCTCTCACCGTATAACCTCAGCACCTCTGGAAGACCT 1800
Db 3399 CCAGCTTACCACCTTTGGGCATTCAGCTCTCACCGTATAACCTCAGCACCTCTGGAAGACCT 3458
QY 1801 GAACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGA 1860
Db 3459 GAACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGA 3518
QY 1861 AGCCACACAGGGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGCTGTGCCAGGTC 1920
Db 3519 AGCCACACAGGGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGCTGTGCCAGGTC 3578
QY 1921 CTGGGAGAGAGCCATCTCGCCAAAACAAAGTGCCTACTATATCAACCCACGAGACTCAAAAC 1980
Db 3579 CTGGGAGAGAGCCATCTCGCCAAAACAAAGTGCCTACTATATCAACCCACGAGACTCAAAAC 3638
QY 1981 AACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAA 2040
Db 3639 AACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAA 3698
QY 2041 T 2041
Db 3699 T 3699

RESULT 7
AAD37230
ID AAD37230 standard; DNA; 4182 BP.
XX
AC AAD37230;
XX
DT 21-AUG-2002 (first entry)
XX
DE Human dystrophin minigene delta4173.
XX

Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin; adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD; Becker muscular dystrophy; ds.

Homo sapiens.

WO200183695-A2.

08-NOV-2001.

27-APR-2001; 2001WO-US13677.

28-APR-2000; 2000US-200777P.

(XIAO/) XIAO X.

Xiao X;

WPI; 2002-049342/06.

New dystrophin minigene for treating Duchenne or Becker muscular dystrophy comprises an N-terminal domain or modified N-terminal domain, rod repeats, H1 and H4 domains and a cysteine rich domain of a dystrophin gene -

Example 1; Page 43-44; 71pp; English.

The present invention relates to an isolated nucleotide sequence encoding a dystrophin minigene. The minigene comprises N-terminal or modified N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control element. The dystrophin minigene in operable linkage with an expression control element, in a recombinant adeno-associated virus or retrovirus is useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is human dystrophin minigene delta4173 containing nucleotides 1-1992 (N-terminus, hinge H1 and rods R1, R2 and R3), 8059-10227 (rods R22, R23 and R24, hinge H4 and CR domain) and 11047-11058 (dystrophin last 3 amino acids).

Sequence 4182 BP; 1309 A; 927 C; 970 G; 976 T; 0 other;

Query Match 83.4%; Score 1702.8; DB 24; Length 4182; Best Local Similarity 88.5%; Pred. No. 0; Matches 1969; Conservative 0; Mismatches 72; Indels 183; Gaps 4;

QY	1	TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAGAAGATATTATCGTGGCTTCTTC	60
DB	1020	TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAGAAGATATTATCGTGGCTTCTTC	1079
QY	61	TGCTGAGGACACATTGCAAGCACAAGGAGAGATTCTTAATGATGTGGAAGTGGTGAAGA	120
DB	1080	TGCTGAGGACACATTGCAAGCACAAGGAGAGATTCTTAATGATGTGGAAGTGGTGAAGA	1139
QY	121	CCAGTTTTCATCTACTGAGGGGTACATGATGGATTGACAGCCCATCAGGGCCGGTTGG	180
DB	1140	CCAGTTTTCATCTACTGAGGGGTACATGATGGATTGACAGCCCATCAGGGCCGGTTGG	1199
QY	181	TAATATTCTACAATTGGGAAGTAAAGCTGATTGGAACAGGAAAAATTATCAGAAGATGAAGA	240
DB	1200	TAATATTCTACAATTGGGAAGTAAAGCTGATTGGAACAGGAAAAATTATCAGAAGATGAAGA	1259
QY	241	AACTGAAGTACAAGAGCAGATGAATCTCTAAATTCAAGATGGGAATGCCTCAGGGTAGC	300
DB	1260	AACTGAAGTACAAGAGCAGATGAATCTCTAAATTCAAGATGGGAATGCCTCAGGGTAGC	1319
QY	301	TAGCATGGAACAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAAACT	360
DB	1320	TAGCATGGAACAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAAACT	1379
QY	361	GAAAGATTGAATGACTGGCTAACAAAAACAGAAAGAAACAAAGGAAAAATGGAGGAAGA	420

Db	1380	GAAAGAGTTGAATGACTGGCTAAACAAAAACAGAAAGAAACAAAGGAAAAATGGAGGAAGA	1439
QY	421	GCCTCTTGGACCTGATCTTGAAGACCTAAAAACGCCAAGTACAACAAACATAAGGTCTTCA	480
Db	1440	GCCTCTTGGACCTGATCTTGAAGACCTAAAAACGCCAAGTACAACAAACATAAGGTCTTCA	1499
QY	481	AGAAGATCTAGAACAAACAAAGTCAAGGTCAATTTCTCTCACTCACATGGTGGTAGT	540
Db	1500	AGAAGATCTAGAACAAACAAAGTCAAGGTCAATTTCTCTCACTCACATGGTGGTAGT	1559
QY	541	TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTTGAAGAACAACTTAAGGTATTGGG	600
Db	1560	TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTTGAAGAACAACTTAAGGTATTGGG	1619
QY	601	AGATCGATGGGCAACATCTGTAGATGGACAGAACCCGCTGGGTCTTTTACAAGACCA	660
Db	1620	AGATCGATGGGCAACATCTGTAGATGGACAGAACCCGCTGGGTCTTTTACAAGACAT	1679
QY	661	GCCTGAC-----CTAGCTCCTGGACTGACCACATATTGGAGCCTCTCTCTACTCA	708
Db	1680	CCTTCTCAAAATGGCAACGCTTACTGAAGAACAGTGCCTTTTACTGTCATGGCTTTTCAGA	1739
QY	709	GACTGTACTCTGGTGACACA-----ACCTGTGGTTACTAAGGAAAACTGCCATCTC	759
Db	1740	AAAAAAGATGCAGTGAACAAGATTACACACAACTGGCTTTAAAGATCAAAAATGAAATGTT	1799
QY	760	-----	759
Db	1800	ATCAAGTCTTCAAAAACTGGCCGTTTAAAGCGGATCTAGAAAAAGAAAAAGCAATCCAT	1859
QY	760	-----CAAACCTAGAAATGCCATCTTCTCTGATGTTGGAG-----	793
Db	1860	GGGCAAACTGTATTCACTCAAAACAAGATCTTCTTCAACACTGAAGAATAAGTCAGTGAC	1919
QY	794	-----	793
Db	1920	CCAGAAGACGGAAGCATGGCTGGATAACTTTGCCCGGTGTTGGGATAATTTAGTCCAAAA	1979
QY	794	-----GTACCTACTCATAGATTACTGCAACAGTTCCTCCCTGGACCTGGA	837
Db	1980	ACTTGAAAAAGAGTACAGCACAGACTCATAGATTACTGCAACAGTTCCTCCCTGGACCTGGA	2039
QY	838	AAAGTTTCTTGCCTGGCTTACAGAAGCTGAAACAACTGCAATGTCTCTACAGGATGCTAC	897
Db	2040	AAAGTTTCTTGCCTGGCTTACAGAAGCTGAAACAACTGCAATGTCTCTACAGGATGCTAC	2099
QY	898	CCGTAAGGAAAGGCTCTAGAAAGACTCCAAGGGAGTAAAGAGCTGATGAAACAAATGGCA	957
Db	2100	CCGTAAGGAAAGGCTCTAGAAAGACTCCAAGGGAGTAAAGAGCTGATGAAACAAATGGCA	2159
QY	958	AGACCTCCAAGGTGAATTTGAAGCTCACACAGATGTTTATCACAACTGGATGAAACACAG	1017
Db	2160	AGACCTCCAAGGTGAATTTGAAGCTCACACAGATGTTTATCACAACTGGATGAAACACAG	2219
QY	1018	CCAAAAATCCTGAGATCCCTGGAAAGGTTCCGATGATGCAGTCTCTCTGAGGAACTTTT	1077
Db	2220	CCAAAAATCCTGAGATCCCTGGAAAGGTTCCGATGATGCAGTCTCTCTGAGGAACTTTT	2279
QY	1078	GGATAACATGAACCTCAAGTGAAGTGAAGTTCGCACTTCTCTGAGGAACTTTTCTGCT	1137
Db	2280	GGATAACATGAACCTCAAGTGAAGTGAAGTTCGCACTTCTCTGAGGAACTTTTCTGCT	2339
QY	1138	TTTGGAAAGCCAGTTCTGACCAGTGAAGCGTCTGCACCTTCTCTGAGGAACTTTTCTG	1197
Db	2340	TTTGGAAAGCCAGTTCTGACCAGTGAAGCGTCTGCACCTTCTCTGAGGAACTTTTCTG	2399
QY	1198	GTGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTGGAGGCGACTTTCC	1257
Db	2400	GTGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTGGAGGCGACTTTCC	2459
QY	1258	AGCAGTTCAGAAGCAGAACGATGTACATAGGGCCTTCAAGAGGGGAATTTGAAAACTAAAGA	1317
Db	2460	AGCAGTTCAGAAGCAGAACGATGTACATAGGGCCTTCAAGAGGGGAATTTGAAAACTAAAGA	2519

QY 1318 ACCTGTAATCATGAGTACTCTTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGA 1377
|||||
Db 2520 ACCTGTAATCATGAGTACTCTTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGA 2579
QY 1378 AGGACTAGAGAAAACCTCTACAGGAGCCCGAGAGAGCTGCCTCTCCTGAGGAGAGAGCCAGAA 1437
|||||
Db 2580 AGGACTAGAGAAAACCTCTACAGGAGCCCGAGAGAGCTGCCTCTCCTGAGGAGAGAGCCAGAA 2639
QY 1438 TGTCACCTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGAAAAATTGAA 1497
|||||
Db 2640 TGTCACCTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGAAAAATTGAA 2699
QY 1498 CTTGCACTCCGCTGACTGGCAGAGAAAATAGATGAGACCTTTGAAAGACTTCCAGGAACT 1557
|||||
Db 2700 CTTGCACTCCGCTGACTGGCAGAGAAAATAGATGAGACCTTTGAAAGACTTCCAGGAACT 2759
QY 1558 TCAAGAGGCCACCGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATC 1617
|||||
Db 2760 TCAAGAGGCCACCGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATC 2819
QY 1618 CTGGCAGCCCGTGGCGATCTCCTCATTTGACTCTCTCCAAAGATCACCTCGAGAAAGTCAA 1677
|||||
Db 2820 CTGGCAGCCCGTGGCGATCTCCTCATTTGACTCTCTCCAAAGATCACCTCGAGAAAGTCAA 2879
QY 1678 GGCACCTTCGAGAGAAAATTGGCCTCTGAAAGAGAACGTCAGCCACGTCATGACCTTGC 1737
|||||
Db 2880 GGCACCTTCGAGAGAAAATTGGCCTCTGAAAGAGAACGTCAGCCACGTCATGACCTTGC 2939
QY 1738 TCGCCAGCTTACCACTTTGGGCAATTCAGCTCTCACCGTATACCTTCAGCACTCTGGAAGA 1797
|||||
Db 2940 TCGCCAGCTTACCACTTTGGGCAATTCAGCTCTCACCGTATACCTTCAGCACTCTGGAAGA 2999
QY 1798 CCTGAACACCATGGAAGCTTCTGCGAGGTGGCGTCTGAGGACCGAGTCAGGCAGCTGCA 1857
|||||
Db 3000 CCTGAACACCATGGAAGCTTCTGCGAGGTGGCGTCTGAGGACCGAGTCAGGCAGCTGCA 3059
QY 1858 TGAAGCCACAGGGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGCTCTGTCCAGGG 1917
|||||
Db 3060 TGAAGCCACAGGGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGCTCTGTCCAGGG 3119
QY 1918 TCCCTGGGAGAGAGCCATCTCGCCAAAAGTGCCTACTATATCAACCCAGAGACTCA 1977
|||||
Db 3120 TCCCTGGGAGAGAGCCATCTCGCCAAAAGTGCCTACTATATCAACCCAGAGACTCA 3179
QY 1978 AACAACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAA 2037
|||||
Db 3180 AACAACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAA 3239
QY 2038 TAAT 2041
|||||
Db 3240 TAAT 3243
|||||
RESULT 8
AD37232
D AAD37232 standard; DNA; 2169 BP.
X
C AAD37232;
T
X
21-AUG-2002 (first entry)
X
Human dystrophin rod, hinge and CR domain regions encoding DNA #1.
E
X
Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
W
X
adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
W
X
Becker muscular dystrophy; ds.
X
Homo sapiens.
S
X
WO200183695-A2.
N
X
08-NOV-2001.
C

XX 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX (XIAO/) XIAO X.
PA
XX
PI Xiao X;
XX
XX WPI; 2002-049342/06.
DR
XX
PT
PT New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -
XX
PS Example 1; Page 45-46; 71pp; English.
XX
CC The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is human
CC dystrophin DNA fragment encoding rods R22, R23 and R24, hinge H4 and
CC CR domain regions.
XX
SQ Sequence 2169 BP; 623 A; 529 C; 524 G; 493 T; 0 other;

Query Match 60.9%; Score 1242; DB 24; Length 2169;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 800 ACTCATAGATTACTGCAACAGTTCCTCCCTGGACCTGGAAAAGTTTCTTGCCTGGCTTACA 859
|||
Db 2 ACTCATAGATTACTGCAACAGTTCCTCCCTGGACCTGGAAAAGTTTCTTGCCTGGCTTACA 61
QY 860 GAAGCTGAAACAACACTGCCAATGTCTACAGGATGCTACCCGTAAGGAAAGGCTCCTAGAA 919
|||
Db 62 GAAGCTGAAACAACACTGCCAATGTCTACAGGATGCTACCCGTAAGGAAAGGCTCCTAGAA 121
QY 920 GACTCCAAGGGAGTAAAGAGCTGATGAAAACAATGGCAAGACCTCCAAGGTGAAATTGAA 979
|||
Db 122 GACTCCAAGGGAGTAAAGAGCTGATGAAAACAATGGCAAGACCTCCAAGGTGAAATTGAA 181
QY 980 GCTCACACAGATGTTTATCACAACCTGGATGAAAACAGCCCAAAAATCTCTGAGATCCCTG 1039
|||
Db 182 GCTCACACAGATGTTTATCACAACCTGGATGAAAACAGCCCAAAAATCTCTGAGATCCCTG 241
QY 1040 GAAGGTTCCGATGATGCTCCTGTTTACAAAGACGTTTGGGATAACATGAACCTTCAAGTGG 1099
|||
Db 242 GAAGGTTCCGATGATGCTCCTGTTTACAAAGACGTTTGGGATAACATGAACCTTCAAGTGG 301
QY 1100 AGTGAACCTCGGAAAAAGTCTCTCAGGAACTTCTGAGGAGCTTGGGATAACATGAACCTTCAAGTGG 1159
|||
Db 302 AGTGAACCTCGGAAAAAGTCTCTCAGGAACTTCTGAGGAGCTTGGGATAACATGAACCTTCAAGTGG 361
QY 1160 TGGAGGCTCTGCACCTTCTCTCAGGAACTTCTGAGGAGCTTGGGATAACATGAACCTTCAAGTGG 1219
|||
Db 362 TGGAGGCTCTGCACCTTCTCTCAGGAACTTCTGAGGAGCTTGGGATAACATGAACCTTCAAGTGG 421
QY 1220 GAATTAAGCCCGCAGGACCTTATGGAGGCGACTTCCAGCAGTTTCCAGAGCAGAACGAT 1279
|||
Db 422 GAATTAAGCCCGCAGGACCTTATGGAGGCGACTTCCAGCAGTTTCCAGAGCAGAACGAT 481
QY 1280 GTACATAGGGCCTTCAAGAGGGGAAATTGAAAACCTAAAGAACCTGTAATCATGAGTACTCTT 1339
|||
Db 482 GTACATAGGGCCTTCAAGAGGGGAAATTGAAAACCTAAAGAACCTGTAATCATGAGTACTCTT 541

QY 1340 GAGACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAGGACTAGAGAACTCTACCAG 1399
Db 542 GAGACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAGGACTAGAGAACTCTACCAG 601
QY 1400 GAGCCAGAGAGCTGCTCTGAGGAGAGAGCCAGAGATGTCACCTCGGCTTCTACGAAAG 1459
Db 602 GAGCCAGAGAGCTGCTCTGAGGAGAGAGCCAGAGATGTCACCTCGGCTTCTACGAAAG 661
QY 1460 CAGGCTGAGGAGTCAATACTAGTGGGAAAAAATGAACCTGCACTCCGCTGACTGGCAG 1519
Db 662 CAGGCTGAGGAGTCAATACTAGTGGGAAAAAATGAACCTGCACTCCGCTGACTGGCAG 721
QY 1520 AGAAAAATAGATGAGACCCCTTGAAGAGACTCCAGGAACCTTCAAGAGGCCACGGATGAGCTG 1579
Db 722 AGAAAAATAGATGAGACCCCTTGAAGAGACTCCAGGAACCTTCAAGAGGCCACGGATGAGCTG 781
QY 1580 GACCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGATCTC 1639
Db 782 GACCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGATCTC 841
QY 1640 CTCAATTGACTCTCTCCAAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAATTGCG 1699
Db 842 CTCAATTGACTCTCTCCAAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAATTGCG 901
QY 1700 CCTCTGAAAGAGAACGTGAGCCACGTCATGACCTGCTGCGCAGCTTACCACCTTTGGGC 1759
Db 902 CCTCTGAAAGAGAACGTGAGCCACGTCATGACCTGCTGCGCAGCTTACCACCTTTGGGC 961
QY 1760 ATTCAGCTCTCACCGTATAACCTCAGCAGCTCTGGAAGACCTGAAACACAGATGGAAGCTT 1819
Db 962 ATTCAGCTCTCACCGTATAACCTCAGCAGCTCTGGAAGACCTGAAACACAGATGGAAGCTT 1021
QY 1820 CTGCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGAAGCCCCACAGGGACTTTGGT 1879
Db 1022 CTGCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGAAGCCCCACAGGGACTTTGGT 1081
QY 1880 CCAGCATCTCAGCACCTTTCTTCCACGCTGTGTCCAGGGTCCCTGGGAGAGAGCCATCTCG 1939
Db 1082 CCAGCATCTCAGCACCTTTCTTCCACGCTGTGTCCAGGGTCCCTGGGAGAGAGCCATCTCG 1141
QY 1940 CCRAACAAAGTCCCTACTATATCAACACAGAGACTCAAAACAACTTGTCTGGGACCATCCC 1999
Db 1142 CCRAACAAAGTCCCTACTATATCAACACAGAGACTCAAAACAACTTGTCTGGGACCATCCC 1201
QY 2000 AAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAAT 2041
Db 1202 AAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAAT 1243

RESULT 9
AAD37238
ID AAD37238 standard; DNA; 3531 BP.
XX
AC AAD37238;
XX
DT 21-AUG-2002 (first entry)
XX
DE Human dystrophin minigene delta3531.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Homo sapiens.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX

(XIAO/) XIAO X.

Xiao X;

WPI: 2002-049342/06.

New dystrophin minigene for treating Duchenne or Becker muscular dystrophy comprises an N-terminal domain or modified N-terminal domain, rod repeats, H1 and H4 domains and a cysteine rich domain of a dystrophin gene.

Example 1; Page 50-51; 71pp; English.

The present invention relates to an isolated nucleotide sequence encoding a dystrophin minigene. The minigene comprises N-terminal or modified N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control element. The dystrophin minigene in operable linkage with an expression control element, in a recombinant adeno-associated virus (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is human dystrophin minigene delta3531 containing nucleotides 1-1341 (N-terminus, hinge H1 and rod R1), 8059-10227 (rods R22, R23 and R24, hinge H4 and CR domain) and 11047-11058 (dystrophin last 3 amino acids).

Sequence 3531 BP; 1071 A; 809 C; 824 G; 827 T; 0 other;

Query Match 60.9%; Score 1242; DB 24; Length 3531;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1242; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 800 ACTCATAGATTACTGCAACAGTTCCCTGGACCTGGAAAGTTTCTTGCCTGGCTTACA 859

Db 1351 ACTCATAGATTACTGCAACAGTTCCCTGGACCTGGAAAGTTTCTTGCCTGGCTTACA 1410

QY 860 GAAGCTGAAACAACTGCCAATGTCTTACAGGATGCTACCCGTAAGGAAAGGCTCCTAGAA 919

Db 1411 GAAGCTGAAACAACTGCCAATGTCTTACAGGATGCTACCCGTAAGGAAAGGCTCCTAGAA 1470

QY 920 GACTCCAAGGGAGTAAAGAGCTGATGAAACAAATGGCAAGACCTCCAAAGTGAATGAA 979

Db 1471 GACTCCAAGGGAGTAAAGAGCTGATGAAACAAATGGCAAGACCTCCAAAGTGAATGAA 1530

QY 980 GCTCACACAGATGTTTATCACAACTTGATGAAACAGCCCAAAATCCTGAGATCCCTG 1039

Db 1531 GCTCACACAGATGTTTATCACAACTTGATGAAACAGCCCAAAATCCTGAGATCCCTG 1590

QY 1040 GAAGGTTCCGATGATGCAGTCTTACAAAGACGTTTGGATAACATGAACATCAAGTGG 1099

Db 1591 GAAGGTTCCGATGATGCAGTCTTACAAAGACGTTTGGATAACATGAACATCAAGTGG 1650

QY 1100 AGTGAACCTCGGAAAAAGTCTCTCAACATTAGTCCCATTTGGAAGCCAGTTCTGACCAG 1159

Db 1651 AGTGAACCTCGGAAAAAGTCTCTCAACATTAGTCCCATTTGGAAGCCAGTTCTGACCAG 1710

QY 1160 TGAAGCGTCTGCACCTTCTCTGCAGGAACCTTCTGGTGGCTACAGCTGAAAGATGAT 1219

Db 1711 TGAAGCGTCTGCACCTTCTCTGCAGGAACCTTCTGGTGGCTACAGCTGAAAGATGAT 1770

QY 1220 GAATTAAGCCGCGCAGCACCTTATTTGAGGCGGACCTTCCAGCAGTTTCCAGAGCAGAT 1279

Db 1771 GAATTAAGCCGCGCAGCACCTTATTTGAGGCGGACCTTCCAGCAGTTTCCAGAGCAGAT 1830

QY 1280 GTACATAGGGCCTTCAAGAGGGAATTTGAAACTAAAGAACCTTCTAATCATGAGTACTCT 1339

Db 1831 GTACATAGGGCCTTCAAGAGGGAATTTGAAACTAAAGAACCTTCTAATCATGAGTACTCT 1890

QY 1340 GAGACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAAGGACTAGAGAAACTCTACCAG 1399

Db 1891 GAGACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAAGGACTAGAGAAACTCTACCAG 1950

2Y 1460 CAGGCTGAGGAGGTCAATACTAGTGGGAAAAATTGAACCTGCACCTCCGCTGACTGGCAG 1519
|||||
Db 2768 CAGGCTGAGGAGGTCAATACTAGTGGGAAAAATTGAACCTGCACCTCCGCTGACTGGCAG 2827
|||||
QY 1520 AGAAAAATAGATGAGACCCCTTGAAGACTCCAGGAACCTTCAAGAGGCCCGCATGAGCTG 1579
|||||
Db 2828 AGAAAAATAGATGAGACCCCTTGAAGACTCCAGGAACCTTCAAGAGGCCCGCATGAGCTG 2887
|||||
QY 1580 GACCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGGATCTC 1639
|||||
Db 2888 GACCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGGATCTC 2947
|||||
QY 1640 CTCATTGACTCTCTCCAAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGGAGAAATTGCG 1699
|||||
Db 2948 CTCATTGACTCTCTCCAAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGGAGAAATTGCG 3007
|||||
QY 1700 CCTCTGAAAGAGAACGCTGAGCCACGTCATGACCTTGCTGCCAGCTTACCACCTTTGGGC 1759
|||||
Db 3008 CCTCTGAAAGAGAACGCTGAGCCACGTCATGACCTTGCTGCCAGCTTACCACCTTTGGGC 3067
|||||
QY 1760 ATTGAGCTCTCACCGTATAACCTCAGCACCTCTGGAGAGACCTGAACACCCAGATGGAAGCTT 1819
|||||
Db 3068 ATTGAGCTCTCACCGTATAACCTCAGCACCTCTGGAGAGACCTGAACACCCAGATGGAAGCTT 3127
|||||
QY 1820 CTGCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGAAAGCCACAGGGACTTTGGT 1879
|||||
Db 3128 CTGCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGAAAGCCACAGGGACTTTGGT 3187
|||||
QY 1880 CCAGCATCTCAGCACCTTCTTCCACAGTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCG 1939
|||||
Db 3188 CCAGCATCTCAGCACCTTCTTCCACAGTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCG 3247
|||||
QY 1940 CCNAACAAAGTGCCCTACTATATCAACACGAGACTCAAAACAACTGCTGGGACCATCCC 1999
|||||
Db 3248 CCNAACAAAGTGCCCTACTATATCAACACGAGACTCAAAACAACTGCTGGGACCATCCC 3307
|||||
QY 2000 AAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAAT 2041
|||||
Db 3308 AAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAAT 3349
|||||

RESULT 11
AAD37240
ID AAD37240 standard; DNA; 3510 BP.
XX
AC AAD37240;
XX
DT 21-AUG-2002 (first entry)
XX
DE Human dystrophin minigene delta3510.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Homo sapiens.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
XX
PI XIAO X;
XX
DR WPI; 2002-049342/06.
XX
PT New dystrophin minigene for treating Duchenne or Becker muscular
dystrophy comprises an N-terminal domain or modified N-terminal domain,

rod repeats, H1 and H4 domains and a cysteine rich domain of a
dystrophin gene -
XX Example 1; Page 51-52; 71pp; English.
PS
XX The present invention relates to an isolated nucleotide sequence encoding
a dystrophin minigene. The minigene comprises N-terminal or modified
N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
domains and cysteine-rich domains of dystrophin or utrophin genes. The
invention also relates to a recombinant adeno-associated virus (AAV)
comprising dystrophin minigene operably linked to an expression control
element. The dystrophin minigene in operable linkage with an expression
control element, in a recombinant adeno-associated virus or retrovirus is
useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
dystrophy (BMD) in a mammalian subject. The present sequence is human
dystrophin minigene delta3510 containing nucleotides 1-1668 (N-terminus,
hinge H1 and rods R1 and R2), 8407-10227 (rods R23 and R24, hinge H4 and
CR domain) and 11047-11058 (dystrophin last 3 amino acids).
XX Sequence 3510 BP; 1073 A; 787 C; 828 G; 822 T; 0 other;
SQ
Query Match 51.6%; Score 1053; DB 24; Length 3510;
Best Local Similarity 76.0%; Pred. No. 1.3e-288;
Matches 1552; Conservative 0; Mismatches 0; Indels 489; Gaps 1;
QY 1 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGATATATCGTGGCTTCTTTC 60
|||||
Db 1020 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGATATATCGTGGCTTCTTTC 1079
|||||
QY 61 TGCTGAGGACACATTGCAAGCACAAAGGAGAGATTCTTAATGATGTGGAAGTGGTAAAGA 120
|||||
Db 1080 TGCTGAGGACACATTGCAAGCACAAAGGAGAGATTCTTAATGATGTGGAAGTGGTAAAGA 1139
|||||
QY 121 CCAGTTTCATCTACTGAGGGGTACATGATGATTTGACAGCCCATCAGGGCCGGTGG 180
|||||
Db 1140 CCAGTTTCATCTACTGAGGGGTACATGATGATTTGACAGCCCATCAGGGCCGGTGG 1199
|||||
QY 181 TAATATTCTACAATTGGGAAGTAAGCTGATTGGAACAGGAAAAATTTACAGAAGATGAAGA 240
|||||
Db 1200 TAATATTCTACAATTGGGAAGTAAGCTGATTGGAACAGGAAAAATTTACAGAAGATGAAGA 1259
|||||
QY 241 AACTGAAGTACAAGACAGATGAATCTCTCTAAATTTCAAGATGGGAATGCCTCAGGGTAGC 300
|||||
Db 1260 AACTGAAGTACAAGACAGATGAATCTCTCTAAATTTCAAGATGGGAATGCCTCAGGGTAGC 1319
|||||
QY 301 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAAACT 360
|||||
Db 1320 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAAACT 1379
|||||
QY 361 GAAAGATTGAATGACTGGCTAACAAAAACAGAAAGAAAGAAACAAAGGAAAAATGGAGGAAGA 420
|||||
Db 1380 GAAAGATTGAATGACTGGCTAACAAAAACAGAAAGAAAGAAACAAAGGAAAAATGGAGGAAGA 1439
|||||
QY 421 GCCTCTTGGACCTGATCTTGAAGACCTAAACCGCAAGTACAACAAACATAAGGTGCTTCA 480
|||||
Db 1440 GCCTCTTGGACCTGATCTTGAAGACCTAAACCGCAAGTACAACAAACATAAGGTGCTTCA 1499
|||||
QY 481 AGAAGATCTAGAACAAAGAACAGTCAGGGTCAATTTCTCACTCACATGGTGGTGGTAGT 540
|||||
Db 1500 AGAAGATCTAGAACAAAGAACAGTCAGGGTCAATTTCTCACTCACATGGTGGTGGTAGT 1559
|||||
QY 541 TGATGAATCTAGTGGAGATCAGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGG 600
|||||
Db 1560 TGATGAATCTAGTGGAGATCAGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGG 1619
|||||
QY 601 AGATCGATGGGCAAAACATCTGTAGATGGACAGAGACCGCTGGTCTTTTACAAGACCA 660
|||||
Db 1620 AGATCGATGGGCAAAACATCTGTAGATGGACAGAGACCGCTGGTCTTTTACAAGCA --- 1676
|||||
QY 661 GCCTGACCTAGCTCCTGGACTGACCACCTATTGGAGCCCTCTCCTACTCAGACTGTTACTCT 720
Db 1677 ----- 1676

QY	721	GGTGACACAACCTGTGGTTACTAAGGAAACTGCCATCTCCAACCTAGAAATGCCATCTTC	780
Db	1677	-----	-----
QY	781	CTTGATGTTGGAGGTACCTACTCATAGATTACTGCAACAGTTCCCCCTGGACCTGGAAAA	840
Db	1677	-----	-----
QY	841	GTTTCTTCCTGGCTTACAGAAAGCTGAAACAACATGCCAATGTCTACAGGATGCTACCCG	900
Db	1677	-----	-----
QY	901	TAAGGAAAGGCTCCTAGAAGACTCCAAGGGAGTAAAGAGCTGATGAACAATGGCAAGA	960
Db	1677	-----	-----
QY	961	CCTCCAAGGTGAATTGAAGCTCACACAGATGTTTATCAACAACCTGGATGAAAACAGCCA	1020
Db	1677	-----	-----
QY	1021	AAAAATCCTGAGATCCCTGGAAGGTTCCGATGATGCAGTCCTGTTACAAAGACGTTTGA	1080
bb	1677	-----	-----
QY	1081	TAACATGAAGTTCAAGTGGAGTGAAGTTCCGAAAAAAGTCTCTCAACATTAGGTCCCATT	1140
bb	1677	-----	-----
Y	1141	GGAAGCCAGTTCTGACCAGTGGAGCGCTCTGCACCTTTCTCTGCAGSNACTTCTGGTGTG	1200
bb	1677	-----	-----
Y	1201	GCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTGGAGGCGACTTCCAGC	1260
b	1731	GCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTGGAGGCGACTTCCAGC	1790
Y	1261	AGTTCAGAAAGCAGAACCGATGTACATAGGGCCTTCAAGAGGGAATTTGAAAACCTAAAGAACC	1320
b	1791	AGTTCAGAAAGCAGAACCGATGTACATAGGGCCTTCAAGAGGGAATTTGAAAACCTAAAGAACC	1850
Y	1321	TGTAATCATGAGTACTCTTGAGACTGTACGATATTTCTGACAGAGCAGCCCTTTGGAAGG	1380
b	1851	TGTAATCATGAGTACTCTTGAGACTGTACGATATTTCTGACAGAGCAGCCCTTTGGAAGG	1910
Y	1381	ACTAGAGAAACTCTACCAGGAGCCGAGAGCTGCCCTCCTGAGGAGAGAGCCCAAGATGT	1440
b	1911	ACTAGAGAAACTCTACCAGGAGCCGAGAGCTGCCCTCCTGAGGAGAGAGCCCAAGATGT	1500
Y	1441	CACCTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTTGAACCT	1560
b	1971	CACCTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTTGAACCT	2030
Y	1501	GCACTCCGCTGACTGGCAGAGAAAAATAGATGAGACCTTTGAAAGACTCCAGGAACCTCA	1560
b	2031	GCACTCCGCTGACTGGCAGAGAAAAATAGATGAGACCTTTGAAAGACTCCAGGAACCTCA	2090
Y	1561	AGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATCCTG	1620
b	2091	AGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATCCTG	2150
Y	1621	GCAGCCCGTGGCGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGC	1680
b	2151	GCAGCCCGTGGCGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGC	2210
Y	1681	ACTTCGAGAGAAATTTGCGCCTCTGAAGAGAAACGTGAGCCACGTCAATGACCTTGCTCG	1740
b	2211	ACTTCGAGAGAAATTTGCGCCTCTGAAGAGAAACGTGAGCCACGTCAATGACCTTGCTCG	2270
Y	1741	CCAGCTTACCACCTTTGGGCATTCAGTCTCACCGTATAACCTCAGCACTCTGGAAGACCT	1800
b	2271	CCAGCTTACCACCTTTGGGCATTCAGTCTCACCGTATAACCTCAGCACTCTGGAAGACCT	2330
Y	1801	GAACACCAGATGGAAGCTTCTGCAGGTGGCCGTGAGGACCGAGTCAGGCAGCTGGATGA	1860

Db	2331		GAACACCAGATGGAAGCTTCTG	AGGTGGCCGTCGAGGACCGAGT	CGAGGACGCTGCATGA	2390
QY	1861		AGCCACACAGGGACTTTGGTCC	CAGCATCTCAGCACTTTCTTTCC	ACGCTGTGTCCAGGGTCC	1920
Db	2391		AGCCACACAGGGACTTTGGTCC	CAGCATCTCAGCACTTTCTTTCC	ACGCTGTGTCCAGGGTCC	2450
QY	1921		CTGGGAGAGAGCCATCTCGCCAA	ACAAAGTGCCCTACTATATCA	AACCAACGAGACTCAAAC	1980
Db	2451		CTGGGAGAGAGCCATCTCGCCAA	ACAAAGTGCCCTACTATATCA	AACCAACGAGACTCAAAC	2510
QY	1981		AACCTGCTGGGACCATCCCAAA	ATGACAGAGCTCTACCAAGTCTT	TAGCTGACCTGAATAA	2040
Db	2511		AACCTGCTGGGACCATCCCAAA	ATGACAGAGCTCTACCAAGTCTT	TAGCTGACCTGAATAA	2570
QY	2041	T	2041			
Db	2571	T	2571			

RESULT 12
AAD37259
ID AAD37259 standard; DNA; 4476 BP.
XX

AC AAD37259;

DT 21-AUG-2002 (first entry)
XX

DE Adeno-associated virus vector plasmid, AAV-MCK-3510.
XX

Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin; adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD; Becker muscular dystrophy; ds.

Chimeric	Homo sapiens.
Chimeric	Unidentified.
OS	
OS	
vv	

WO200183695-A2.

08-NOV-2001.

27-APR-2001; 2001WO-US13677.

28-APR-2000; 2000US-200777P.

(XIAO/) XIAO X.

Xiao X;

WPI; 2002-049342/06.

New dystrophin minigene for treating Duchenne or Becker muscular dystrophy comprises an N-terminal domain or modified N-terminal domain, rod repeats, H1 and H4 domains and a cysteine rich domain of a dystrophin gene -

Example 1; Page 63-65; 71pp; English.

The present invention relates to an isolated nucleotide sequence encoding a dystrophin minigene. The minigene comprises N-terminal or modified N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control element. The dystrophin minigene in operable linkage with an expression control element, in a recombinant adeno-associated virus with an expression control element, in a recombinant adeno-associated virus or retrovirus is useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is AAV vector plasmid construct containing human dystrophin minigenes, a muscle creatine kinase (MCK) promoter and a small polyA signal sequence.

Sequence 4476 BP; 1252 A; 1096 C; 1127 G; 1001 T; 0 other;

Mon Apr 28 09:28:35 2003

Query Match									
Best Local Similarity 51.6%; Score 1053; DB 24; Length 4476;									
Matches 1552; Conservative 0; Mismatches 0; Indels 489; Gaps 1;									
2Y	1	TGAAGTAAACCTGGACCGTTATCAACAGCCTTTAGAGAAAGTATTATCGTGGCTTCTTTC	60						
Db	1776	TGAAGTAAACCTGGACCGTTATCAACAGCCTTTAGAGAAAGTATTATCGTGGCTTCTTTC	1835						
QY	61	TGCTGAGGACACATTGCAAGCACAAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA	120						
Db	1836	TGCTGAGGACACATTGCAAGCACAAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA	1895						
QY	121	CCAGTTTCATCTACTCATGAGGGGTACATGATGGATTGACAGCCCATCAGGCCGGGTTGG	180						
Db	1896	CCAGTTTCATCTACTCATGAGGGGTACATGATGGATTGACAGCCCATCAGGCCGGGTTGG	1955						
QY	181	TAATATTCTACAAATGGGAAGTAAGCTGATTTGGAACAGGAAATTTATCAGAAAGATGAAGA	240						
Db	1956	TAATATTCTACAAATGGGAAGTAAGCTGATTTGGAACAGGAAATTTATCAGAAAGATGAAGA	2015						
QY	241	AACCTGAAGTACACAGACAGATGAATCTCCTAAATCAAGATGGGAATGCCTCAGGGTAGC	300						
Db	2016	AACCTGAAGTACACAGACAGATGAATCTCCTAAATCAAGATGGGAATGCCTCAGGGTAGC	2075						
QY	301	TAGCATGGAAAAACAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT	360						
Db	2076	TAGCATGGAAAAACAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT	2135						
QY	361	GAAAGAGTTGAATGACTGGCTTAACAAAAACAGAAAGAAAGAAATGGAGGAAGA	420						
Db	2136	GAAAGAGTTGAATGACTGGCTTAACAAAAACAGAAAGAAAGAAATGGAGGAAGA	2195						
QY	421	GCCTCTTGGACCTGATCTTGAAGACCTAAACGCCAAGTACACAAACATAAAGGTCTTCA	480						
Db	2196	GCCTCTTGGACCTGATCTTGAAGACCTAAACGCCAAGTACACAAACATAAAGGTCTTCA	2255						
QY	481	AGAAGATCTAGAACAAGAAACAAGTCAGGGTCAATTTCTCACTCACATGGTGGTGTAGT	540						
Db	2256	AGAAGATCTAGAACAAGAAACAAGTCAGGGTCAATTTCTCACTCACATGGTGGTGTAGT	2315						
QY	541	TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTGGAAAGAACAACTTAAGGTATTGGG	600						
Db	2316	TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTGGAAAGAACAACTTAAGGTATTGGG	2375						
QY	601	AGATCGATGGGCAACATCTGTAGATGGACAGAGACCGCTGGGTTCTTTTACAAGA---	2432						
Db	2376	AGATCGATGGGCAACATCTGTAGATGGACAGAGACCGCTGGGTTCTTTTACAAGA---	2432						
QY	661	GCCTGACCTAGCTCCTGGACTGACCACTATTGGAGCCCTCTCCTACTCAGACTGTTACTCT	720						
Db	2433	-----	2432						
QY	721	GGTGACACAACCTGTGGTTACTAAGGAAACTGCCATCTCCAAACTAGAAAATGCCATCTTC	780						
Db	2433	-----	2432						
QY	781	CTTGATGTTGGAGGTACCTACTCATAGATTACTGCAACAGTTCCCCCTGGACCTGGAAAA	840						
Db	2433	-----	2432						
QY	841	GTTTCTTGCTGGCTTACAGAACTGCAACAACTGCCAATGTCTTACAGGATGCTACCCG	900						
Db	2433	-----	2432						
QY	901	TAAGGAAAGGCTCCTAGAAAGTCCAAAGGAGTAAAGAGCTGATGAAACAATGGGAAGA	960						
Db	2433	-----	2432						
QY	961	CCTCCAAGGTGAAATTAAGCTCACACAGATGTTTATCAACAACCTGGATGAAACACAGCCA	1020						
Db	2433	-----	2432						
QY	1021	AAAAATCCTGAGATCCCTGGAGGTTCCGATGATGCAGTCTCCTGTACAAAAGACGTTTGGGA	1080						

Db	2433	-----	2432
QY	1081	TAACATGAACCTTCAAGTGGAGTGAACCTTCGGAAAAAGTCTCTCAACATTAGGTCCCATTT	1140
Db	2433	-----	2432
QY	1141	GGAAGCCAGTTCTGACCAGTGAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGTG	1200
Db	2433	-----	2432
QY	1201	GCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTGAGGCGGACTTTCCAGC	1260
Db	2487	GCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTGAGGCGGACTTTCCAGC	2546
QY	1261	AGTTTCAGAGCAGAAACGATGTACATAGGCGCTTCAAGAGGGAATTTGAAAACTAAAAAACC	1320
Db	2547	AGTTTCAGAGCAGAAACGATGTACATAGGCGCTTCAAGAGGGAATTTGAAAACTAAAAAACC	2606
QY	1321	TGTAATCATGAGTACTCTTGAGACTGTACGAATATTCTGACAGAGCAGCCTTTGGAAGG	1380
Db	2607	TGTAATCATGAGTACTCTTGAGACTGTACGAATATTCTGACAGAGCAGCCTTTGGAAGG	2666
QY	1381	ACTAGAGAAACTCTACAGAGAGCCCGAGAGAGCTGCCTCCTGAGGAGAGAGCCCAAGATGT	1440
Db	2667	ACTAGAGAAACTCTACAGAGAGCCCGAGAGAGCTGCCTCCTGAGGAGAGAGCCCAAGATGT	2726
QY	1441	CACCTGGCTTCTACGAAAGCAGCTGAGGAGGTCAATACTGAGTGGGAAAAATTTGAACCT	1500
Db	2727	CACCTGGCTTCTACGAAAGCAGCTGAGGAGGTCAATACTGAGTGGGAAAAATTTGAACCT	2786
QY	1501	GCACCTCCGCTGACTGGCAGAGAAATAGATGAGACCCCTTGAAGACTCCAGGAACCTTCA	1560
Db	2787	GCACCTCCGCTGACTGGCAGAGAAATAGATGAGACCCCTTGAAGACTCCAGGAACCTTCA	2846
QY	1561	AGAGGCCACGGATGAGCTCAAGCTCGCCCAAGCTGAGGTGATCAAGGGATCCTG	1620
Db	2847	AGAGGCCACGGATGAGCTCAAGCTCGCCCAAGCTGAGGTGATCAAGGGATCCTG	2906
QY	1621	GCAGCCCGTGGCGATCTCCTCATTTGACTCTCTCAAGATCACCTCGAGAAAGTCAAGGC	1680
Db	2907	GCAGCCCGTGGCGATCTCCTCATTTGACTCTCTCAAGATCACCTCGAGAAAGTCAAGGC	2966
QY	1681	ACTTCGAGGAGAAATTCGCGCTCTGAAAGAGAACGTCAGCCACGTCATGACCTTGCCTCG	1740
Db	2967	ACTTCGAGGAGAAATTCGCGCTCTGAAAGAGAACGTCAGCCACGTCATGACCTTGCCTCG	3026
QY	1741	CCAGCTTACCACTTTGGGCAATTCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCT	1800
Db	3027	CCAGCTTACCACTTTGGGCAATTCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCT	3086
QY	1801	GAACACCAAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGA	1860
Db	3087	GAACACCAAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGA	3146
QY	1861	AGCCACACAGGAGCTTTGGTCCAGCATCTCAGCACTTTCTTCCACGTCGTCCAGGGTCC	1920
Db	3147	AGCCACACAGGAGCTTTGGTCCAGCATCTCAGCACTTTCTTCCACGTCGTCCAGGGTCC	3206
QY	1921	CTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCACGAGACTCAAAC	1980
Db	3207	CTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCACGAGACTCAAAC	3266
QY	1981	AACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACAGTCTTTAGCTGACCTGAATAA	2040
Db	3267	AACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACAGTCTTTAGCTGACCTGAATAA	3326
QY	2041	T 2041	
Db	3327	T 3327	

RESULT 13

*

AAD37241
 ID AAD37241 standard; DNA; 1821 BP.
 XX AC AAD37241;
 XX DT 21-AUG-2002 (first entry)
 XX DE Human dystrophin rod, hinge and CR domain regions encoding DNA #2.
 XX KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
 KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
 XX Becker muscular dystrophy; ds.
 OS Homo sapiens.
 XX PN WO200183695-A2.
 XX PD 08-NOV-2001.
 XX PF 27-APR-2001; 2001WO-US13677.
 XX PR 28-APR-2000; 2000US-200777P.
 XX PA (XIAO/) XIAO X.
 XX PI xiao X;
 XX WPI; 2002-049342/06.
 XX
 XX New dystrophin minigene for treating Duchenne or Becker muscular
 PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
 PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
 PT dystrophin gene -
 XX
 XX Example 1; Page 52-53; 71pp; English.
 XX
 CC The present invention relates to an isolated nucleotide sequence encoding
 CC a dystrophin minigene. The minigene comprises N-terminal or modified
 CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
 CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
 CC invention also relates to a recombinant adeno-associated virus (AAV)
 CC comprising dystrophin minigene operably linked to an expression control
 CC element. The dystrophin minigene in operable linkage with an expression
 CC control element, in a recombinant adeno-associated virus or retrovirus is
 CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
 CC dystrophy (BMD) in a mammalian subject. The present sequence is human
 CC dystrophin DNA fragment encoding rods R23 and R24, hinge H4 and CR
 CC domain regions.
 XX
 SQ Sequence 1821 BP; 506 A; 451 C; 447 G; 417 T; 0 other;
 Query Match 43.9%; Score 895; DB 24; Length 1821;
 Best Local Similarity 100.0%; Pred. No. 8.1e-244;
 Matches 895; Conservative 0; Mismatches 0; Indels 0; Gaps 0
 QY 1147 CAGTTCTGACCAAGTGAAGCGTCTGCACCTTCTCTGCAGGAACCTTCTGGTGTGGCTACA 1206
 dbb 1 CAGTTCTGACCAAGTGAAGCGTCTGCACCTTCTCTGCAGGAACCTTCTGGTGTGGCTACA 1206
 QY 1207 GCTGAAAGATGATGAATTAGCGGCGAGGCACCTTTCAGAGGGGAATTGAAACCTAAAGAACCTGTAAT 180
 bb 61 GCTGAAAGATGATGAATTAGCGGCGAGGCACCTTTCAGAGGGGAATTGAAACCTAAAGAACCTGTAAT 180
 PY 1267 GAAGCAGAACGATGATACATAGGGCCCTTCAAGAGGGGAATTGAAACCTAAAGAACCTGTAAT 1326
 b 121 GAAGCAGAACGATGATACATAGGGCCCTTCAAGAGGGGAATTGAAACCTAAAGAACCTGTAAT 180
 Y 1327 CATGAGTACTCTTGAGACTGTACGAATATTTCTGCACAGAGCAGCCCTTTGGAAGGACTAGA 1386
 b 181 CATGAGTACTCTTGAGACTGTACGAATATTTCTGCACAGAGCAGCCCTTTGGAAGGACTAGA 1386
 Y 1387 GAACTCTACCAGGAGCCCGAGAGCTGCCTCCTGTAGGAGAGAGCCCGAGAGCTGCTACTCG 1446
 Y 1446 GAACTCTACCAGGAGCCCGAGAGCTGCCTCCTGTAGGAGAGAGCCCGAGAGCTGCTACTCG 1446

1

Db 2077 TAGCATGGAAAAACAAAGCAATTACATAGAGTTTTAAATGGATCTCCAGAATCAGAAACT 2136
QY 361 GAAAGAGTTGAATGACTGGCTAAACAAAACAGAGAAAGAAACAAAGGAAAATGGAGGAAGA 420
Db 2137 GAAAGAGTTGAATGACTGGCTAAACAAAACAGAGAAAGAAACAAAGGAAAATGGAGGAAGA 2196
QY 421 GCCTCTTGGACCTGATCTTGAAGACCTAAACGCCCAAGTACAACAACATAAAGGTGCTTCA 480
Db 2197 GCCTCTTGGACCTGATCTTGAAGACCTAAACGCCCAAGTACAACAACATAAAGGTGCTTCA 2256
QY 481 AGAAGATCTAGAACAAAGAAAGTCAAGGTCAATTCTCTCACTCACATGGTGGTAGT 540
Db 2257 AGAAGATCTAGAACAAAGAAAGTCAAGGTCAATTCTCTCACTCACATGGTGGTAGT 540
QY 541 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAGAAACAACTTAAGGTATTGGG 2316
Db 2317 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAGAAACAACTTAAGGTATTGGG 600
QY 601 AGATCGATGGGCAACATCTGTAGATGGACAGAGACCGCTGGGTCTTTTACAAGACCA 660
Db 2377 AGATCGATGGGCAACATCTGTAGATGGACAGAGACCGCTGGGTCTTTTACAAGACAT 2436
QY 661 GCCTGACCTAGCTCCTGGACTGAC 584
Db 2437 CCTTCTCAAAATGGCAACGTCTTAC 2460

Search completed: April 25, 2003, 08:40:29
Job time : 361.763 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 07:58:12 ; Search time 1925.12 Seconds
(without alignments)
17170.339 Million cell updates/sec

Title: US-09-845-416-6_COPY_1020_3060
Perfect score: 2041
Sequence: 1 tgaagtaaacctggacggtt.....cttagctgacctgaataat 2041

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308016

Minimum DB seq length: 0
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *
1: em_estba: *
2: em_esthum: *
3: em_estin: *
4: em_estmu: *
5: em_estov: *
6: em_estpl: *
7: em_estro: *
8: em_htc: *
9: gb_est1: *
10: gb_est2: *
11: gb_htc: *
12: gb_est3: *
13: gb_est4: *
14: gb_est5: *
15: em_estfun: *
16: em_estom: *
17: gb_gss: *
18: em_gss_hum: *
19: em_gss_inv: *
20: em_gss_pln: *
21: em_gss_vrt: *
22: em_gss_fun: *
23: em_gss_mam: *
24: em_gss_mus: *
25: em_gss_other: *
26: em_gss_pro: *
27: em_gss_rod: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	591.8	29.0	834	13	BI729851
2	491.4	24.1	750	13	BI730168
3	486.6	23.8	579	9	AL121550
4	424.2	20.8	663	13	BM488464
5	394.6	19.3	1490	11	BC009242
6	262.8	12.9	784	12	BG212445

7	260.8	12.8	2334	11	BC011062
8	252	12.3	772	13	BI250598
9	246	12.1	835	13	BI553820
10	241.6	11.8	696	13	BJ075057
11	226	11.1	250	12	BF963618
12	209.4	10.3	502	9	AL602076
13	182.4	8.9	645	10	AW467977
14	176.4	8.6	479	14	BQ304046
15	167.2	8.2	646	13	BI289102
16	164.6	8.1	600	13	BI988528
17	153.4	7.5	481	10	BE557463
18	153.2	7.5	885	11	AK020881
19	150.6	7.4	515	17	AZ780914
20	144.2	7.1	392	10	AW948794
21	142.2	7.0	426	14	BQ319056
22	140.6	6.9	649	14	BQ209933
23	140	6.9	515	10	BB651473
24	133.2	6.5	521	17	AZ226551
25	130.2	6.4	663	10	BB083824
26	122.4	6.0	358	12	BF772373
27	119	5.8	427	12	BE817979
28	118.6	5.8	346	12	BF772375
29	115.8	5.7	504	13	BI289382
30	113.8	5.6	697	10	AV704460
31	107.4	5.3	449	12	BF522118
32	99.6	4.9	317	12	BE815203
33	97.4	4.8	555	9	AI789349
34	96.8	4.7	410	14	BQ349936
35	87.6	4.3	467	9	AI827863
36	86.4	4.2	267	14	BQ379136
37	83.6	4.1	555	10	AW494075
38	81.8	4.0	386	9	AI713038
39	78	3.8	772	13	BM291527
40	77.6	3.8	501	13	BI061271
41	77.2	3.8	589	17	AQ260288
42	77	3.8	434	14	BQ375536
43	76	3.7	423	9	AA460476
44	76	3.7	541	13	BI546771
45	76	3.7	757	9	AL556247

ALIGNMENTS

RESULT 1
BI729851
LOCUS
DEFINITION
60334951f1 NIH_MGC_94 Mus musculus CDNA clone IMAGE:5357162 5',
mRNA sequence.
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
BI729851.1 GI:15706864
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 834)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM1908 row: e column: 03
High quality sequence stop: 796.
Location/Qualifiers
1..834

FEATURES
source

LOCUS BI730168 750 bp mRNA linear EST 20-SEP-2001
DEFINITION 603349711F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5357187 5',
mRNA sequence.
ACCESSION BI730168
VERSION BI730168.1 GI:15707181
KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 750)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: The Cepko Laboratory
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM11908 row: f column: 04
High quality sequence stop: 747.
FEATURES
Location/Qualifiers
1..750
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5357187"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 240 a 161 c 178 g 170 t 1 others
ORIGIN
Query Match 24.1%; Score 491.4; DB 13; Length 750;
Best Local Similarity 88.4%; Pred. No. 3e-108;
Matches 602; Conservative 0; Mismatches 71; Indels 8; Gaps 6;
QY 784 GATGTTGGAGGTACCTACTCATAGATTACTGCAACAGTTCCTCCCTGGACCTGGAAAAAGTT 843
Db 69 GCTGCTTGGGACAGAACTCATAGATTACTGCAGCAGTTCCTCTGGACCTGGAGAGTT 128
QY 844 TCTTGCCCTGG-CTTACAGAAAGCTGAAAC--AACTGCCAATGTCTCTACAGGATGCTACCCG 900
Db 129 TCTTCTGCTGATTTACGGAAGCAGAAACAGAGCTGCCAATGTCTCTACAGGACGCTTCCC 188
QY 901 -TAAGGAAAGGCTCCTAGAAAGACTCCAAAGG--AGTAAAGAGCTGATGAAACAATGGCA 957
Db 189 CTAAGGAGAAAGCTCCTAGCAAGACTCCAGGGGCGAGTCAGAGAGCTGATGAAACCATGGCA 248
QY 958 AGACCTCCAAGGTGAAA-TTGAAGCTCACACAGATGTTTATCACAACTGGATGAAACA 1016
Db 249 AGATCTCCAAGGAGAAAGTTGAAACTCACACAGATATCTATCATCAATCTTGATGAAATG 308
QY 1017 GCCAAAAAATCCTGAGATCCCTGGAAGGTTCCGATGATGATGATGATGATGATGATGATGAT 1076
Db 309 GCCAAAAAATCCTGAGATCCCTGGAAGGTTCCGATGATGATGATGATGATGATGATGATGAT 368
QY 1077 TGGATAACATGAACCTCAAGTGGAGTGAAGTTCGGGAAAAAGTCTCTCAACATTAGGTCCC 1136
Db 369 TGGATAACATGAATTTCAAGTGGAGTGAAGTTCAGAAAAAGTCTCTCAACATTAGGTCCC 428
QY 1137 ATTTGGAAGCCAGTTCTGACCAGTGGAAAGCTCTGCACCTTTCTCTGCAGGAACCTTCTGG 1196
Db 429 ATTTGGAAGCAAGTTCTGACCAGTGGAAAGCTTTGCACTCTTTCTCTCAGGAACCTTCTTG 488

/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5357162"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."
BASE COUNT 275 a 173 c 199 g 187 t
ORIGIN
Query Match 29.0%; Score 591.8; DB 13; Length 834;
Best Local Similarity 90.5%; Pred. No. 1.5e-132;
Matches 665; Conservative 0; Mismatches 67; Indels 3; Gaps 3;
QY 793 GGTACCTACTCATAGATTACTGCAACAGTTCCTCCCTGGACCTGGAAAAAGTTTCTTGCCCTG 852
Db 75 GGAAGAAACTCATAGATTACTGCAGCAGTTCCTCTGGACCTGGAGAAAGTTTCTTCTCCTG 134
QY 853 GCTTACAGAAAGCTGAAACAACCTGCCAATGTCTCTACAGGATGCTACCCGTAAGGAAAGGCT 912
Db 135 GATTACGGAAGCAGAAACAACCTGCCAATGTCTCTACAGGACGCTTCCCGTAAGGAGAGCT 194
QY 913 CCTAGAAAGACTCCAAGGGAGTAAAGAGAGCTGATGAAACAATGGCAAGACTTCCAAGGTGA 972
Db 195 CCTAGAAAGACTCC-AGGGAGTCAGAGAGCTGATGAAACCATGGCAAGACTTCCAAGGAGA 253
QY 973 AATTGAAGCTCACACAGATGTTTATCAACACCTGGATGAAACAGCCCAAAAAATCCTGAG 1032
Db 254 AATTGAAGCTCACACAGATATCTATCAACATCTTGATGAAATGGCCAAAAATCCTGAG 313
QY 1033 ATCCCTGGAAGGTTCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1092
Db 314 ATCCCTGGAAGGTTCCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 373
QY 1093 CAAGTGGAGTGAACCTTCGGAAGAAAGTCTCTCAACATTAGTCCCATTTGGAAAGCAAGTTC 1152
Db 374 CAAGTGGAGTGAACCTTCAGAAAAAGTCTCTCAACATTAGTCCCATTTGGAAAGCAAGTTC 433
QY 1153 TGACCAGTGGAAAGCTGTCACCTTTCTCTGCAAGAACTTCTGGTGTGGCTACAGCTGAA 1212
Db 434 TGACCAGTGGAAAGCTGTCACCTTTCTCTTCAAGAACTTCTTGTGTTGGCTACAGCTGAA 493
QY 1213 AGATGATGAATTAAGCCGGCAGGCAACCTATTGGAGGCGACTTTCAGCAGCTTCAAGAGCA 1272
Db 494 AGATGATGAACCTGAGCCGTCAGGCACCCATCGGTGATGATTTCCAGCAGCTTCAAGAGCA 553
QY 1273 GAACGATGTACATAGGCGCTTCAAGAGGGAATTTGAAACTTAAAGAACTGTAATCATGAG 1332
Db 554 GAATGATATACATAGGCGCTTCAAGAGGGAATTTGAAACTTAAAGAACTGTAATCATGAG 613
QY 1333 TACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAAGGACTAGAGAAACT 1392
Db 614 TACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAAGGACTAGAGAAACT 673
QY 1393 C-TACCAGGAGCCAGAGAGCTGCCTCTCTGAGGAGAGAGCCCAAGAAATGTCACTCGGCTTC 1451
Db 674 CTTACCAGGAGCCAGAGAGCTGCCTCTCTGAAAGAAAGAGCTCAGAAATGTCACTCGGCTTC 733
QY 1452 TACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTAACCTGCACCTC-CGCT 1510
Db 734 TACGAAAGCAGGCTGAGGAGGTCAACGCTGAATGGGACCAATTAACCTGCGCTCAAGCT 793
QY 1511 GACTGGCAGAGAAAA 1525
Db 794 GATTGGCAGAGAAAA 808
RESULT 2
BI730168

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1490)
Strausberg, R.
Direct Submission
Submitted (06-JUN-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
contact: amadan@systemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
Greene, Mark Kettman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 10 Row: j Column: 10
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 5032280
This clone has the following problem: incomplete processing.

Location/Qualifiers
1..1490

/organism="Homo sapiens"
/db_xref="LocusID:1756"
/db_xref="taxon:9606"
/clone="IMAGE:3029414"
/tissue_type="Muscle, rhabdomyosarcoma"
/clone_lib="NIH_MGC_17"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

BASE COUNT 505 a 299 c 328 g 358 t

ORIGIN

Query Match 19.3%; Score 394.6; DB 11; Length 1490;
Best Local Similarity 99.0%; Pred. No. 1.le-84;
Matches 397; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAGAAGTATATCGTGGCTTCTTC 60
Db 1089 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAGAAGTATATCGTGGCTTCTTC 1148
QY 61 TGCTGAGGACACATTGCAAGCACAAAGGAGAGATTCTAATGATGTGGAAGTGGTGAAGA 120
Db 1149 TGCTGAGGACACATTGCAAGCACAAAGGAGAGATTCTAATGATGTGGAAGTGGTGAAGA 1208
QY 121 CCAGTTTCATCTCATGAGGGGTACATGATGGATTGACAGCCCATCAGGCGCGGTGG 180
Db 1209 CCAGTTTCATCTCATGAGGGGTACATGATGGATTGACAGCCCATCAGGCGCGGTGG 1268
QY 181 TAATATTCTACAATTGGGAAGTAAGCTGATTGGAACAGGAAAAATTATCAGAAGATGAAGA 240
Db 1269 TAATATTCTACAATTGGGAAGTAAGCTGATTGGAACAGGAAAAATTATCAGAAGATGAAGA 1328
QY 241 AACTGAAGTACAAGACAGATGAATCTCCTAAATTCAGATGGGAATGCCTCAGGTAGC 300
Db 1329 AACTGAAGTACAAGACAGATGAATCTCCTAAATTCAGATGGGAATGCCTCAGGTAGC 1388
QY 301 TAGCATGGAAGAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAAACT 360
Db 1389 TAGCATGGAAGAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAAACT 1448
QY 361 GAAAGAGTTGATGACTGGCTAACAAAAACAGAGAAAGAA 401
Db 1449 GAAAGAGTTGATGACTGGCTAACAAAAACAGAGAAAGAA 1489

/lab_host="E. coli EMDH10B"
/note="Vector: pCMVSPORT6; Library made from equivalent
pools of total RNA isolated from each tissue (embryonic
muscle 33.3%; juvenile muscle 33.3%; and epiphyseal growth
plate 33.3% of the final RNA pool). Single pass sequencing
from 5'-end"

BASE COUNT 209 a 132 c 172 g 144 t 6 others

ORIGIN

Query Match 20.8%; Score 424.2; DB 13; Length 663;
Best Local Similarity 78.1%; Pred. No. 5.6e-92;
Matches 518; Conservative 0; Mismatches 144; Indels 1; Gaps 1;

QY 940 GCTGATGAAACAATGGCAAGACCTCCAAAGGTGAAATTGAAGCTCACACAGATGTTTATCA 999
Db 1 GCTCATGAAGCAGTGGCAGGATCTACAGGCAGAAATTTGATGCACATACTGACATCT-TNN 59
QY 1000 CAACCTGGATGAAACAGCCAAATAATCCTGAGATCCCTGGAAAGTTCGGATGATGCAGT 1059
Db 60 CAACCTGGATGAAACAGCCAAATAATCCTGAGATCCCTGGAAAGTTCGGATGATGCAGT 119
QY 1060 CCTGTTACAAAGACGTTTGGATAACATGAACCTCAAGTGGAGTGAACCTCGGAAAAAGTC 1119
Db 120 CCTGTTGACAGACGCTCTGGATAACATGAACCTTCAGATGGAGTGAGCTTAGAAGAAATC 179
QY 1120 TCTCAACATAGGTCCCATTTTGAAGCCAGTTCTGACCAAGTGGAAAGCTCTGCACCTTTC 1179
Db 180 TCTAAACATTAGATCTCATTTTGAAGCCAGCAGACAGACCAGTGGAAAGCTTTTACATCTCTC 239
QY 1180 TCTGCAGGAACCTTCTGGTGTGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACC 1239
Db 240 TCTTTCAGGAACCTTTCGGATGGCTGCAATTGAAGGAGGATGAATTAAGCAGCAGCACC 299
QY 1240 TATTGGAGGCGACTTTCAGCAGCTTCAGAAAGCAGAACGATGTACATAGGGCCCTTCAAGAG 1299
Db 300 CATTTGGAGATATTCCTCACTGTGCAGAACGACAGATGATTTTATAGGACTTTTCAAGAG 359
QY 1300 GGAATTGAAACCTAAAGAACCTGTAAATCATGAGTACTCTTGAGACTGTACGAATATTTCT 1359
Db 360 GGAGCTGAAACAAAGAACCTGTTATCATGAATGCACCTTGAGACTGTGCGACTCTTCTCT 419
QY 1360 GACAGAGCAGCCTTTTGGAGGACTAGAGAAACTCTACCAGGAGCCCAAGAGAGCTGCCTCC 1419
Db 420 GGCAGATCAACCAAGTAGAGGACTGGAAAGGTTCTATCCAGAACCAAGAGACCTATCACC 479
QY 1420 TGAGGAGAGAGCCAGAAATGTCACCTGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATAC 1479
Db 480 TGAGGAGAGGCGCCAGAAATGTCACCTGGCTTCTCCGAAAGCAGGACAGATGATGTCAGNAC 539
QY 1480 TGAGTGGGAAAAAATTGAACCTGCACTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCT 1539
Db 540 TGAGTGGGATAAGCTAAATCTACGTTCTGCTGATTGGCAAAGAAAGATAGATGATGCTCT 599
QY 1540 TGAAGACTCCAGGAACCTTCAAGAGGCGCACGGATGAGCTGGACCTCAAGCTGCGCCCAAGC 1599
Db 600 TGAAGACTGCAGGGTCTTTCAGGAGGCAATGGATGAACCTNNNCCTGAAACTGCGCCAGGC 659
QY 1600 TGA 1602
Db 660 TGA 662

RESULT 5
BC009242
LOCUS
DEFINITION
Homo sapiens, dystrophin (muscular dystrophy, Duchenne and Becker
types), includes DXS142, DXS164, DXS206, DXS230, DXS239, DXS268,
DXS269, DXS270, DXS272, clone IMAGE:3029414, mRNA.

ACCESSION BC009242
VERSION BC009242.1 GI:14714379
KEYWORDS HTC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens

[illegible]

QY	2	GAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAGAAGTATTATATCGTGGCTCTTCT	61
Db	44	GACATGGATTGGACAGCTACCAGATAGCGCTAGAGGAAGTGTGACGTGGCTGCTGTCC	103
QY	62	GCTGAGGACACATTTGCAAGCACACAGGAGAGATTTCTAATGATGTGGAAAGTGTGAAAGAC	121
Db	104	GCGGAGGACACGTTCCAGGAGCAAGATGACATTTCTGATGATGTGCGAAGAAGTCAAAGAG	163
QY	122	CAGTTTCATACTCATGAGGGGTACATGATGGATTTCACAGCCCATCAGGGCCGGTTGGT	181
Db	164	CAGTTGC-TACCCATGAAACTTTTATGATGGAGCTGACAGCACACCGAGAGCAGCGTGGG	222
QY	182	AATATTCTACAATTTGGGAAGTAAGCTGATTGGAACAGGAAATATCAGAAGATGAAGAA	241
Db	223	AGCGTCTGCAGGCTGGCAACCAGCTGATGACACAAGGACTCTGTTCAGAGGAGGAGGAC	283
QY	242	ACTGAAGTACAAGAGCAGATGAATCTCCTAAATCAAGATGGGAATGCCTCAGGGTAGCT	301
Db	283	GTGAGATCCAGGAACAGATGACCTTGCTGAATGCAAGGTGGGAGCGCTCCGGGTGGAG	344
QY	302	AGCATGGAAAAACAAAGCAATTTACATAGAGTTTTAATGGATCTCCAGAATCAGAAAAC	366
Db	343	AGCATGGAGAGGACGCTCCCGGCTGCACGACGCTCTGTATGGAGTGCAGAAGAACAACAGCTG	401
QY	362	AAAGAGTTGAATGACTGGCTAAACAAAAACAGAAAGAAAGAACAGGAAATGGAGGAAGAG	422
Db	403	CAGCAGCTCTCAAGCTGSGCTGGCCCTCACAGAAGAGCGCATTCAGAAGATGGAGAGCCTC	466
QY	422	CCTCTTGGACCTGATCTTGAAGACCTAAAAACGCCAAGTACAACAACATAAGGTCTTCAA	488
Db	463	CCGCTGGGTGATGACCTGCCCTCCCTGCAGAAGCTGCTTCAAGAACAATAAATTTTGC	522
QY	482	GAAGATCTAGAACAAAGACAAAGTCAGGGTCAATTTCTCTCACTCACATGGTGGTAGTT	544
Db	523	AATGACCTTGAAGCTGAACAGGTGAAGGTAAATTCCTTAACCTCACATGGTGGTAGTTGTG	588
QY	542	GATGAATCTAGTGGAGATCACGCACTGCTGCTTTTGGAAAGAACAACTTAAGGTAAATGGGA	600
Db	583	GATGAAAAACAGTGGGGAGAGTGCCACAGCTCTTCTTGGAAAGATCAGTTTACAGAAACTGGGT	644

Simpson, A.J.
shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genom
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?t1=QV2&t2=QV2-NN004>)
211200-570-d02&t3=2000-12-21&t4=1)
Seq primer: puc 18 forward
High quality sequence stop: 249.

FEATURES
SOURCE

```
Location/Qualifiers
1. .250
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NN0045"
/dev_stage="Adult"
```

```

/dev_stage="Adult"
/note="Organ: nervous_normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

```

86 a	59 c	57 g	48 t
------	------	------	------

[illegible]

Query Match	11.1%;	Score 226;	DB 12;	Length 250;
Best Local Similarity	100.0%;	Pred. No. 4.4e-44;		
Matches 226;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

814 GCAACAGTTCCCCCTGGACCTGGAAAAGTTCTTGCCTTACAGAGCTGAACAAC 873

Db	5	GCAACAGTTCCCCCTGGACCTGGAAAAAGTTTCTTGCCCTGCGCATTACAGAAGACTTAATTCAGCATC
QY	874	TGCCAATGTCTCTACAGGATGCTACCCGTAGSAAAGGCTCCTAGAAGACTCCAAGGAGT

Db 65 TGCCAAATGTCCTACAGGATGCTACCCGTAAGGAAAGGCTCCTAGAAAGACTCCCAAGGAGT 124
024 AAAAAAGAGCTGATGAAACCAATGGCAAGACCTCCACGGTGAATTTGAAGCTCACACAGATGT 993

[illegible]

Qy 994 TTATCACAACTGGATGAAAAACAGCCAAAAAATCCTGAGATCCCTG 1039
 |||||
 db 185 TTATCACAACTGGATGAAAAACAGCCAAAAAATCCTGAGATCCCTG 230

RESULT 12
AT.602076

LOCUS	AL602076	502 bp	mRNA	linear	EST I4-AUG-2000
DEFINITION	DKFZp313B029_r1	313 (synonym: hlcc2)	Homo sapiens	cdna clone	
FEATURES	DKFZp313B029_5'	mRNA sequence.			

AL602076	AL602076.1	GI:15165582
ACCESSION		
VERSION		

VERSION
KEYWORDS
SOURCE
EST.
human.

SOURCE	Hommat.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa;
	Chordata; Vertebrata; Euteleostomi;
	Catarrhini; Hominiidae; Homo.

REFERENCE
1 (bases 1 to 502)
AUTHORS Ansoorge,W., Wirkner,U., Mewes,W., Weil,B. and Wiemann,S.
TITLE EST (Ansoorge,W., Wirkner,U., Mewes,H.W., Weil,B. and Wiemann,S.)
JOURNAL Unpublished (1999)
COMMENT Contact: Ansoorge W

```
library"
/tissue_type="whole embryo"
/dev_stage="stage 25"
/note="Vector: pBSRN3; site_1: NotI; site_2: EcoRI; cDNAs
were oligo-dT primed and directionally cloned. Staging
according to Nieuwkoop and Faber. Library is subtracted
and was constructed by N. Garrett and A.M. Zorn,
(Wellcome/CRC Institute)."
224 a 160 c 152 g 160 t
```

Query Match	11.8%;	Score 241.6;	DB 13;	Length 696;
Best Local Similarity	64.5%;	Pred. No. 9.9e-48;		
Matches 361. Conservative	0;	Mismatches 199;	Indels 0;	Gaps 0;

[illegible]

db 16 AGTGGACAGCCTGAACTCTCAGGCCAGCAGTTGGCAGAAGCAAGTGGACAAGACCTTGG 75

2y 1542 AAAGACTCCAGGAACCTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGGCCCAAGCTG 1601

76 AGAAGCTGAAGGAGTTACAAAGTCGAATGGAAGAACTGGATGTGGAGTTAATGAGTGAG 135

[illegible]

QY
Db

1662 ACCTCGAGAAAGTCAAGGCAC TTCGAGGAGAAAATTGGCCCTCTGAAAGAGAACGTGAGCC
|| ||| | - ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
196 ACATTTGAGACAACAACAGCATTTGGAGAGGAAATTTGCCCAGTCAGCTCTAAGAGTACAAA

172

1722 ACGTCAATGACCTTGCTGCCAGCTTACCACATTGGGCATTGAGCTCTCACCGTATAACC 178

Db	256	CACTGAATGATATGCCAGTCAGCTTTGTACCTTTGACATTCAGCCATCTGCATTAACAT	311
OY	1782	TCAGCACTCTGGAAGACCTGAACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACC	184

Db 316 CTCGCCAGTTGGATGACCTGAACATCAGATGGAAGCTTTACAGGCAGCTGTTGAAGAAC 375
1842 CAGTCAGCCAGCTGCATGAAGCCACAGGGGACTTTGGTCCAGCATCTCAGCACCTTCTTT 190

DQ
184 Z GAGCAGGCGCTGCATCAACCCGCCGGTGTTCCTTTTGGAACCTCTCCCACACTTTCTCT

DB
376 GTCTCAAACAACCTCCAAGAAGCACATCGGGATT TTGGACCCTGCTCCCCAACACTTTCTCT

Qy	1902	CCACGTCGTGCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAACAACAAAGTGCCCTACTATA	1968
Db	436	CTACTTCAGTTCAGCTTCATGGCAGCGGTCAAGTATCACTCAACAAAAGTACCCTACTACA	495

[illegible]

DB	496	TCAACCATCAAAACACTACTCTGTTGGGATCACCCAAAAAAGACAGAGACATTAATGAGAT
QY <th>2022</th> <th>CTTTAGCTGACCTGAATAAT 2041</th>	2022	CTTTAGCTGACCTGAATAAT 2041

Db 556 CTCTAGGTGACTTAAATAAT 575

RESULT 11
BF963618
LOCUS
BF963618
250 bp
mRNA linear
Homo sapiens CDNA,
NN0045
EST 22-JAN-2001
mRNA sequence.

DEFINITION QV2-NN0045-211200-570-8
ACCESSION BF963618
VERSION BF963618.1 GI:12380893

KEYWORDS	EST.
SOURCE	human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Catarrhini; Hominoidea; Homo.

REFERENCE
AUTHORS

1 (bases 1 to 250)

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Mammalla, E., Lucena, L., Bordin, S., Costa, F.F., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., deOliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza S.J. and

MIPS

Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis,
Research Center (DKFZ); Email s.wiemann@dkfz-
heidelberg.de; sequenced by EMBL (European Molecular Biology
Laboratories, Heidelberg/Germany) within the cDNA
sequencing consortium of the German Genome Project.

Query Match 10.3%; Score 209.4; DB 9; Length 502;
Best Local Similarity 69.3%; Pred. No. 5.7e-40;
Matches 285; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

1691 GAAATTGCGCCTCTGAAAGAGAACGTGAGCCACGTCAATGACCTTGCTGCCAGCTTACC 1750
||||||| | | || | ||||| | | ||||| |
61 GAAATTGCACCAATCAACTTTAAAGTTAAAACGGTGAATGATTTATCCAGTCAGCTGTCT 120

1751 ACTTTGGGCATTGACGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCTGAACACCAGA 1810
121 CCACITGACCTGCATCCCTCTCTAAAGATGTCTGCCAGCTAGATGACCTTAATATGCCA 180

1811 TGAAGCTTCTGCAGGTGGCCGTCAGACCGAGTCAGGCAGCTGCATGAAGCCCACAGG 1870
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
181 TGGAAACTTTACAGGTTTTCCGTGGATGATCGCCTTAACAGCTTCAGGAAGCCCACAGA 240

1871 GACITTTGGTCCAGCATCTCAGCACTTTCTTTCCAGCTGTGTCAGGGTCCCTGGGAGAGA 1930
|| ||||| ||| | ||||| ||||| || ||||| ||||| || ||||| ||||| || ||||| |||||
241 GATTTTGGACCATCCTCTCAGCATTTTCTCTACGTGAGTCCAGTGCCCGTGGCAAAGA 300

1931 GCCATCTCGCCCAACAAAGTGCCTACTATATCAACCACGAGACTCAAACAACITTGCTGG 1990
||||| || | || ||||| || ||||| || ||||| || ||||| || ||||| || |||||
301 TCCATTTCACATAATAAAGTGCCTATTACATCAACCATCAAACGCAGACCACCTGTGG 360

1991 GACCATCCAAAATGACAGAGCTCTACCAGTCTT^{AGCTGACCTGAATAAT} 2041
 ||||| ||||| || ||| || ||||| ||||| |||||
 361 GACCATCTAAAATGACCGAAGCTTT^{CAATCCCTGCTGACCTGAATAAT} 411

SULT 13
467977/C
CUS
AW467977 645 bp mrna linear EST 24-FEB-2000
he29g08.x1 NCI_CGAP_CML1 Homo sapiens cDNA clone IMAGE:2920478 3'
similar to SW:UTRO_HUMAN P46939 UTROPHIN ; , mrna sequence.
AW467977
SESSION

JOURNAL
COMMENT

Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Elisabeth Paietta, Jonathan D. Licht, M.D.,
Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: Life
Technologies, Inc. CDNA Library Arrayed by: Christa Prange, The
I.M.A.G.E. Consortium DNA Sequencing by: Washington University
Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
Seq primer: -40UP from Gibco
High quality sequence stop: 419.

QY	899	CGTAAGGAAAGGCTCTCTAGAAGACTCCAAGGGAGTAAAGAGCTGATGAAACAATGGCAA	958
Db	585	CATCGGGAGAATGCTCTTCAGGATAGTATCTTGGCCAGGGAACCTCAAACAGCAGATGCAG	526

QY	959	GACCTCAAGTGAAATTGAAGCTCACACAGATGTTTATCACAACTGGATGAAAACAGC	101
Db	525	GACATCCAGGCAGAAATTGATGCCCAATGACATATTTAAAGGCATTTACGGGAAACAGG	466

[illegible]

QY	1079	GATAACATGAACTTCAAGTGGAGTGAACTTCGGAAAAAAGTCTCTCAACAATTAGGTCCCAT - - -	1138
Db	405	GATGATATGAACCAAAAGATGGAATGACTTAAAAGCAA AATCTGCTAGCATCAGGGGCCAT 	346

Qy	1139	TTGGAAGCCAGTCTGTACCAGTGAAGCGTCTGCACCTTTCTCTGCAGGAAC	TCTGGTG	1198
Db	345	TTGGAGGCCAGCGCTGAGAAGTGAACAGGTTGCTGATGTCCTTAGAAGAACTGATCAA	A	286

Qy	1199	GGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGCACCTATTGGAGCGACTT	TCCA	1258
		- -		
Db	285	GGCTGAATATGAAAGATGAAGAGCTTAAGAAACAATGCTATTGGAGGAGAT	TCCA	226

QY	1259	GCAGTTCAGAAGCAGAACGATGTACATAGGGCCCTTCAAGAGGGAATTGAAAACATAAAGAA	1318
Db	225	GCCTTACAGCTCCAGTATGACCATGTGAAGGCCCTGAGACGGGAGTTAAAGGAGAAAGAA	166

QY 1319 CCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTCTGACAGAGCAGCCTTT 1374
Db 165 TATTCTGCTCGAATGCTGTCGACCAGCCCGAGTTTTCTTGGCTGATCAGCCCAATTGAG 106

QY 1375 ----GGAAGGACTAGAGAACTCTACCAGGAGCCACAGAGAGCTGCCTCCTGAGGAGAGA 1429
||||| - - - - -
Db 105 GCCCCTGAAGAGCCAAGAAGAACCTACAATCAAAAACAGAAATTAACTCCTGAGGAGAGA 46

QY 1776 ATAACCTCAGCACTCTGGAAGACCTTCTGCAAGTGGCGCCGTCG 1835

FEATURES source

```

/dev_stage="ADULT"
/lab_host="DH10B (Life Technologies)"
/note="vector: pT7T3D-pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; The UI-R-DK0
library is a subtracted library derived from a mixture of
five individually tagged normalized rat libraries:
brain-nRBP (20%), heart-nRHP (20%), kidney-nRKP (20%),
aorta-nRAP (20%), and placenta-nRPP (20%). Each original
library was constructed from a mixture of equal amounts of

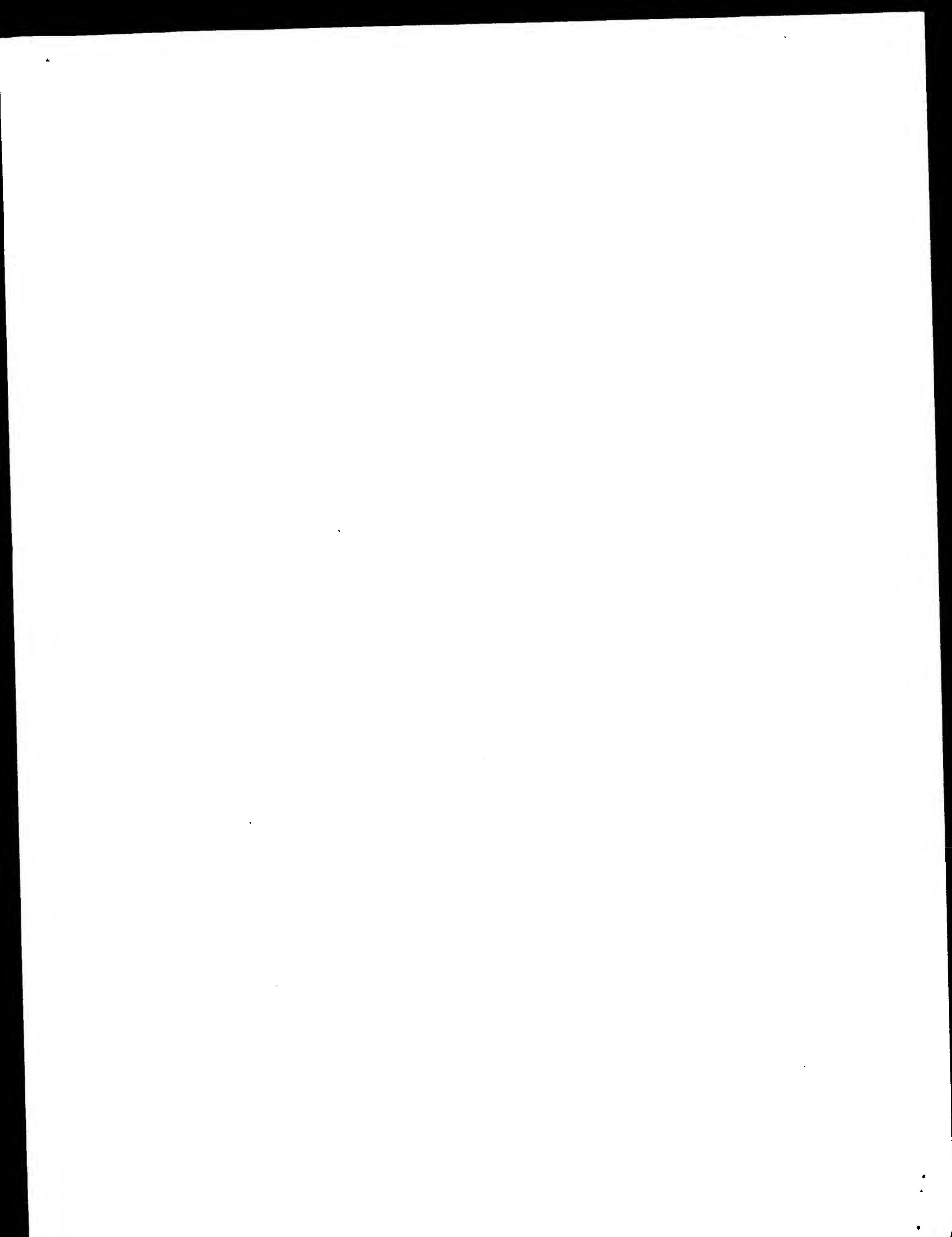
```


RNA from seven different developmental time-points: embryonic day 17, embryonic day 19, embryonic day 21, adult day 1, adult day 12, adult day 75, and adult day 200. (Exception: the aorta pool does not contain embryonic day 17 RNA and the placenta pool contains only the three embryonic stages). Each library was normalized individually according to the procedure described by Bonaldo, Lennon & Soares (Genome Research Genome 6: 791-806, 1996). For construction of the DK0 subtracted library, plasmid DNA from each of the five individually tagged, normalized libraries was mixed in the proportions specified above and electroporated into competent bacteria for production of single-stranded circular DNA representing the pool of libraries. Single-stranded circular DNA representing these five normalized libraries was then used as a tracer in a subtractive hybridization with a driver (PCR amplified inserts from a plasmid DNA template preparation) comprising: a) a set of about 1,000 arrayed clones from each of the five non-normalized libraries of brain (CT0s), heart (CS0s), kidney (CU0s), aorta (CW0s), and placenta (CX0s). The resulting pool of approximately 5,000 clones represented about 33.3% of the final driver population. A set of about 2,000 arrayed clones from each of the five normalized libraries of brain (CT0), heart (CS0), kidney (CU0), aorta (CW0), and placenta (CX0). The resulting pool of about 10,000 clones represented about 66.6% of the final driver population. TAG_LIB=UI-R-DK0

TAG_TISSUE=rat placenta pool
TAG_SEQ=TCACGACAGT

BASE COUNT 131 a 187 c 129 g 198 t 1 others
ORIGIN

Query Match		8.2%	Score 167.2;	DB 13;	Length 646;
Best Local Similarity		60.1%	Pred. No. le-29;		
Matches 277;		Conservative	0;	Mismatches 184;	Indels 0; Gaps 0;
QY	197	GGAAGTAAGCTGATTGGACACAGGAAATATCAGAAGATGAAGAACTGAACTACAAAG	256		
Db	639	GGCAACCAACTGATGACACAAAGGCACTCTGTCGGATGAGGAGGAGTGTGATTCAGGAG	580		
QY	257	CAGATGAATCTCCTAAATTCAGATGGGAATGCCTCAGGGTAGCTAGTAACTGAAAGAGTTGAATGAC	316		
Db	579	CAGATGACCTTGCTGAACGCCAGATGGGAGGCGCTCCGGGTAGAGAGCATGGAGAGACAG	520		
QY	317	AGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAACTGAAAGAGTTGAATGAC	376		
Db	519	TCCCGGTTGCATGATGCTCTGATGGAACCTGCAGAAAGAACAGCTGCAGCAGCTCTCAGGC	460		
QY	377	TGGCTAACAAACACAGAAAGAAACAGGAAATGGAGGAGAGCCCTCTTGGACCTGAT	436		
Db	459	TGGCTGACCTTCACAGAGGAGCGCATTCAGAAGATGGAAGAGCTCCAGTGGGTGACGAC	400		
QY	437	CTTGAAGACCTAAACGCCCAAGTACAACACATAGGTGCTTCAAGAAGATCTAGAACAA	496		
Db	399	CTACCCCTCCCTGCAAAACCTGCTTGAAGAACATATAAAGTTTGCAAGTGACCTCGAAGCT	340		
QY	497	GAACAAGTCAGGTCAATCTCTCACTACATGGTGGTGGTAGTTGATGAATCTAGTGA	556		
Db	339	GAGCAGGTGAAGGTGAATTCCTTAACATATGTTGGTGGTGGTGGTGGTGGTGGTGGG	280		
QY	557	GATCAGCAACTGCTGCTTTGGAGAACAACTTAAGGTATTGGGAGATCGATGGGCAAAAC	616		
Db	279	GAGAGCGCCACAGCTGTTTGGAGAGATCAGTTACAGAAACTGGGTGAGCGCTGGACAGCT	220		
QY	617	ATCTGTAGATGGACAGAGACCGCTGGGTCTTTTACAAGA	657		
Db	219	GTATGCCGCTGGACTGAAGAACCTTGGAAACAGGTTGCAAGA	179		



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 08:10:12 ; Search time 46.4852 Seconds
(without alignments)
13465.096 Million cell updates/sec

Title: US-09-845-416-6_COPY_1020_3060
Perfect score: 2041
Sequence: 1 tgaagtaaacctggaccgtt.....cttagctgacctgaataat 2041

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 875442

Minimum DB seq length: 0
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents_NA:*
1: /cgn2-6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2-6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2-6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2-6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2-6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2-6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query Length	ID	Description
1	40.4	2.0	2223	1	US-08-257-073-4
2	38.6	1.9	1995	1	US-08-425-069-3
3	38.6	1.9	1995	2	US-08-317-844B-3
4	38.2	1.9	1131	6	5180810-3
5	38.2	1.9	1784	6	5180810-2
6	38	1.9	1394	4	US-09-247-155-76
7	36.8	1.8	1886	6	5210183-1
8	36.6	1.8	1845	4	US-08-887-534A-22
9	36.2	1.8	289	4	US-09-007-005-17
10	36.2	1.8	289	4	US-09-244-796-17
11	36.2	1.8	2447	2	US-09-014-969-14
12	36	1.8	2873	4	US-08-630-915A-193
13	35.8	1.8	790	4	US-09-461-474-13
14	35.6	1.7	4868	1	US-08-139-937-12
15	35.6	1.7	4868	5	PCT-US93-11310-12
16	35.2	1.7	3466	1	US-08-468-036-38
17	35.2	1.7	3466	2	US-08-376-843-38
18	35	1.7	2169	4	US-09-434-408-3
19	35	1.7	4766	5	PCT-US93-07261-10
20	34.8	1.7	2442	4	US-09-040-485-1
21	34.8	1.7	3337	1	US-08-072-610-1
22	34.8	1.7	3337	2	US-08-719-822B-1
23	34.8	1.7	3337	4	US-09-092-458-1
24	34.6	1.7	198	5	PCT-US95-10668-3
25	34.6	1.7	198	5	PCT-US95-10668-4
26	34.6	1.7	2277	1	US-08-676-967-5
27	34.6	1.7	2277	1	US-08-676-974-5

C 28	34.6	1.7	2277	2	US-09-098-487-5	Sequence 5, Appli
C 29	34.6	1.7	2704	4	US-08-857-076-44	Sequence 44, Appl
C 30	34.6	1.7	3499	4	US-08-857-076-43	Sequence 43, Appl
31	34.4	1.7	717	4	US-09-461-697-189	Sequence 189, App
32	34.4	1.7	774	4	US-09-461-697-187	Sequence 187, App
33	34.4	1.7	819	4	US-09-461-697-185	Sequence 185, App
34	34.4	1.7	1047	4	US-09-671-950-1	Sequence 1, Appli
35	34.4	1.7	1047	4	US-09-671-950-3	Sequence 3, Appli
36	34.4	1.7	1047	4	US-09-671-950-5	Sequence 5, Appli
37	34.4	1.7	1047	4	US-09-671-950-7	Sequence 7, Appli
38	34.4	1.7	1047	4	US-09-671-950-9	Sequence 9, Appli
39	34.4	1.7	1047	4	US-09-671-950-11	Sequence 11, Appli
40	34.4	1.7	1047	4	US-09-671-950-13	Sequence 13, Appli
41	34.4	1.7	1669	4	US-09-461-697-184	Sequence 184, App
C 42	34.2	1.7	621	4	US-09-328-111-204	Sequence 204, App
C 43	34.2	1.7	734	3	US-09-014-583-1	Sequence 1, Appli
44	34.2	1.7	750	4	US-08-961-527-370	Sequence 370, App
45	34.2	1.7	1089	3	US-08-642-807A-27	Sequence 27, Appl

ALIGNMENTS

RESULT 1
US-08-257-073-4
; Sequence 4, Application US/08257073
; Patent No. 5766597
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: de Taisne, Charles
; APPLICANT: Tine, John A.
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: New York
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/257,073
FILING DATE: 09-JUN-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/075,783
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/852,305
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/672,183
FILING DATE: 20-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2570
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066 CURTMS
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2223 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-257-073-4


```

;
; LENGTH: 1354
;
; TYPE: DNA
;
; ORGANISM: Homo sapiens
;
; FEATURE:
;
; NAME/KEY: CDS

```


Mon Apr 28 09:28:36 2003

```

; LOCATION: 46..675
; FEATURE:
; NAME/KEY: sig_peptide
; LOCATION: 46..87
; OTHER INFORMATION: Von Heijne matrix
; OTHER INFORMATION: score 5.9
; OTHER INFORMATION: seq LTLGLSLILAGL/IV
; FEATURE:
; NAME/KEY: polyA_signal
; LOCATION: 1363..1368
; FEATURE:
; NAME/KEY: polyA_site
; LOCATION: 1382..1394
; US-09-247-155-76

Query Match      1.9%; Score 38; DB 4; Length 1394;
Best Local Similarity 59.1%; Pred. No. 0.17;
Matches 65; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 309 AAAACAAAGCAATTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACTGAAAGAGT 368
   ||||| || || || || || || || || || || || || || || || || || || ||
Db 1087 AAAACAATTTAAATAATTTCAAATTTTAAATTTGCTGCACCATAAGATGAATAAGAGC 1028

QY 369 TGAATGACTGGCTAACAAAAACAGAAAGAAAGAAAGAAAGAAATGGAGGAA 418
   || || || || || || || || || || || || || || || || || || || || ||
Db 1027 TTACTTAAAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 978

RESULT 7
5210183-1
; Patent No. 5210183
; APPLICANT: LINDAHL, GUNNAR; FRITHZ, ELISABET; HEDEN, LARS-OLOF
; TITLE OF INVENTION: PROTEIN ARP, WITH IMMUNOGLOBULIN A
; BINDING ACTIVITY, THE CORRESPONDING VECTORS AND HOSTS, REAGENT
; KIT AND PHARMACEUTICAL COMPOSITION
; NUMBER OF SEQUENCES: 3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/270,099
; FILING DATE: 14-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 186,097
; FILING DATE: 25-APR-1988
; SEQ ID NO: 1:
; LENGTH: 1886
; 5210183-1
```

```

Query Match      1.8%; Score 36.8; DB 6; Length 1886;
Best Local Similarity 45.5%; Pred. No. 0.5;
Matches 131; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 214 AACAGGAAAATTTATCAGAAAGATGAAGAACTGAAGTACAAGAGCAGATGAATCTCTAAA 273
   || || || || || || || || || || || || || || || || || || || || ||
Db 261 AGCGTGAAGTGGCTTAAAGAAATATAACGCGTTACTTAAGGAAATGAGGAGCTCAAGGT 320

QY 274 TTCAAGATGGGAATGCCCTCAGGTTAGCTAGCATGGAAAAACAAAGCAATTTACATAGAGT 333
   || || || || || || || || || || || || || || || || || || || || ||
Db 321 AGAAGCTGAAAAATATCTATCTTATGCTGACGATAAAGAAAAAGATCTCTCAATCTAGAGC 380

QY 334 TTTAATGGATCTCCAGAACTCAGAAAGTGAATGACTGGCTAAGCAAAAAACAGA 393
   ||||| || || || || || || || || || || || || || || || || || || ||
Db 381 ATTAATGGTGAAAATCAAGATCTTCCGAAAAGAGAGGGACAATATCAGGACAAAAATAGA 440

QY 394 AGAAAGAACAAAGGAAAATGGAGGAAGAGCTCTTGGACCTGATCTTTGAAGACCTAAAACG 453
   ||||| || || || || || || || || || || || || || || || || || || ||
Db 441 AGAACTTGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGCAATA 500

QY 454 CCAAGTACAAACATAAAGGTGCTTCAAGAAGATCTAGAACAAAGAA 501
   || || || || || || || || || || || || || || || || || || || || ||
Db 501 TCAATAGAGCAGATGAAGCATTTATCAAGAACAAACAAAGAAACATCA 548
```

RESULT 8
US-08-887-534A-22

```

; Sequence 22, Application US/08887534A
; Patent No. 6455323
; GENERAL INFORMATION:
; APPLICANT: Holden, David W.
; TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/887,534A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 28341/33996
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: (312) 474-6600
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1845 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA (genomic) (p14c15)"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..788
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 856..1842
; US-08-887-534A-22
```

```

Query Match      1.8%; Score 36.6; DB 4; Length 1845;
Best Local Similarity 48.0%; Pred. No. 0.57;
Matches 135; Conservative 0; Mismatches 144; Indels 2; Gaps 1;

QY 150 TGGATTTGACAGCCCATCAGGGCCGGTGGTAATATCTACAATTTGGGAAGTAAGCTGA 209
   ||||| || || || || || || || || || || || || || || || || || || ||
Db 349 TGGCTTTGGATAAACATATATGATGATGTAACAATTAGTTATCCCTACCTGACAATCAGC 408

QY 210 TTGGAACAGGAAAATTTATCAGAAAGATGAAGAACTGAAGTACAAGAGCAGATGAATCTCC 269
   || || || || || || || || || || || || || || || || || || || || ||
Db 409 TACTTACCCAACTATTTACGGAACCTGAAGAAAGGTGCTGTAATGTTGATATGATTTCAC 468

QY 270 TAAATCAAGATGGGAATGCCCTCAGGTAGCTAGCATGGAAAAACAAAGCAATTTACATA 329
   || || || || || || || || || || || || || || || || || || || || ||
Db 469 AAATCGTCAACTTGGATGGGTACAACTATCTTCACGATTAAAGATAGTGTGATTTCATC 528

QY 330 GAGTTTAAATGGATCTCCAGAAATCAGAAACTGAAAGAGTTGAATGACTGGCTAAGCAAAA 389
   ||||| || || || || || || || || || || || || || || || || || || ||
Db 529 AAATTTCTATGATCTT--GAAACATTAAGAAATCAATATGAAGCATTAGCTTATAAAAT 586

QY 390 CAGAAGAAAGAAAGAAAGAAATGGAGGAAGAGCTCTTGGGA 430
   || || || || || || || || || || || || || || || || || || || || ||
Db 587 CAATGAGCATTATGTCAAAAATTTCAATTAATTTGGCTCAGGCA 627
```

RESULT 9
US-09-007-005-17


```

RESULT 11
US-09-014-969-14
; Sequence 14, Application US/09014969
; Patent No. 5965397
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Racie, Lisa A.
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/014,969
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2447 base pairs
; TYPE: nucleic acid

```

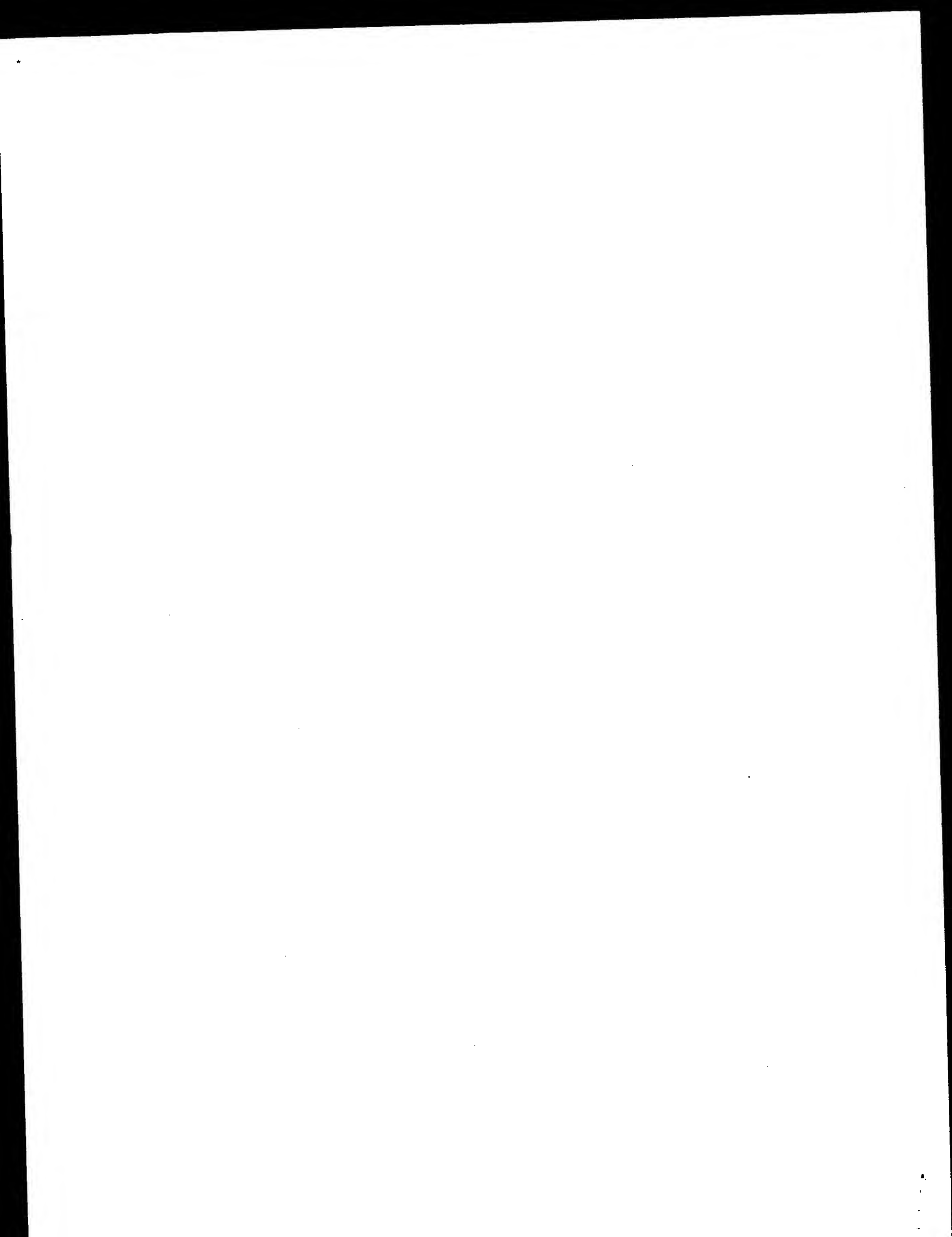

SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/139,937
FILING DATE: 20-OCT-1993
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/979,156
FILING DATE: 20-NOV-1992
ATTORNEY/AGENT INFORMATION:
NAME: CAMPBELL, CATHRYN
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: P-CJ 9370
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-535-9001
TELEFAX: 619-535-8949
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 4868 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-139-937-12

Query Match
Best Local Similarity 1.7%; Score 35.6; DB 1; Length 4868;
Matches 80; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
QY 384 CAAAACAGAGAAAGAAACAAAGGAAATGGAGGAAGAGCCCTCTTGGACCTGATCTTGAAG 443
Db 1921 CAGAAAAGAAATCTGACAAATGAATTACAAAAAGAGCAAGAGCGGAATATCTGAATTAG 1980
QY 444 ACCTAAACGCCCAAGTACAACAACATAAGGTGCTTCAAGAAGATCTAGAACAAAGCAAG 503
Db 1981 AAATAATAATTCATCATTTGAAAATATTTTGAAGAAAGAGCAAGAGCAAGAAAGTACAGA 2040
QY 504 TCAGGGTCAATTCTCTCACTCACATGGTGGTGGT 537
Db 2041 TGAAGAAAAATCAAGCACTGCCATGGAGATGCT 2074

RESULT 15
PCT-US93-11310-12
Sequence 12, Application PC/TUS9311310
GENERAL INFORMATION:
APPLICANT: BOARD OF REGENTS OF THE UNIVERSITY OF TEXAS SYSTEM
TITLE OF INVENTION: CELLULAR GENES ENCODING
TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: CAMPBELL AND FLORES
STREET: 4370 LA JOLLA VILLAGE DRIVE
CITY: SAN DIEGO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 92122
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/11310
FILING DATE: 19-NOV-1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: CAMPBELL, CATHRYN
REGISTRATION NUMBER: 31,815
REFERENCE/DOCKET NUMBER: FP-CJ 9790
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-535-9001
TELEFAX: 619-535-8949
INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:
LENGTH: 4868 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
PCT-US93-11310-12
Query Match
Best Local Similarity 1.7%; Score 35.6; DB 5; Length 4868;
Matches 80; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
QY 384 CAAAACAGAGAAAGAAACAAAGGAAATGGAGGAAGAGCCCTCTTGGACCTGATCTTGAAG 443
Db 1921 CAGAAAAGAAATCTGACAAATGAATTACAAAAAGAGCAAGAGCGGAATATCTGAATTAG 1980
QY 444 ACCTAAACGCCCAAGTACAACAACATAAGGTGCTTCAAGAAGATCTAGAACAAAGCAAG 503
Db 1981 AAATAATAATTCATCATTTGAAAATATTTTGAAGAAAGAGCAAGAGCAAGAAAGTACAGA 2040
QY 504 TCAGGGTCAATTCTCTCACTCACATGGTGGTGGT 537
Db 2041 TGAAGAAAAATCAAGCACTGCCATGGAGATGCT 2074

Search completed: April 25, 2003, 17:43:03
Job time : 76.4852 secs



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 17:39:49 ; Search time 135.61 Seconds
(without alignments)
16376.941 Million cell updates/sec

Title: US-09-845-416-6_COPY_1020_3060
Perfect score: 2041
Sequence: 1 tgaagtaaacctggaccgtt.....cttagctgacctgaataat 2041

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 709820 seqs, 544064369 residues

Total number of hits satisfying chosen parameters: 1410746

Minimum DB seq length: 0
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	153.4	7.5	467	10	US-09-864-761-11083
2	151	7.4	151	10	US-09-864-761-27715
3	140.2	6.9	256	10	US-09-864-761-21956
4	124	6.1	466	10	US-09-864-761-6092
5	43	2.1	436	10	US-09-960-352-10742
6	42.2	2.1	423	10	US-09-960-352-4010
7	40.6	2.0	425	10	US-09-960-352-12395
8	40.4	2.0	428	10	US-09-864-864-204
9	39.8	2.0	418	10	US-09-960-352-12395
10	39.6	1.9	1381	9	US-09-822-846-343
11	39.6	1.9	1594	9	US-10-174-590-183
12	39.6	1.9	1594	9	US-10-176-758-183
13	39.6	1.9	1594	9	US-10-173-706-183
14	39.6	1.9	1594	9	US-10-175-738-183
15	39.6	1.9	1594	9	US-10-175-738-183
16	39.6	1.9	1594	9	US-10-175-752-183
17	39.6	1.9	1594	9	US-10-176-482-183
18	39.6	1.9	1594	9	US-10-176-757-183
19	39.6	1.9	1594	9	US-10-176-913-183

C 20	39.6	1.9	1594	9	US-10-180-552-183	Sequence 183, App
C 21	39.6	1.9	1594	9	US-10-180-557-183	Sequence 183, App
C 22	39.6	1.9	1594	9	US-10-173-700-183	Sequence 183, App
C 23	39.6	1.9	1594	9	US-10-174-572-183	Sequence 183, App
C 24	39.6	1.9	1594	9	US-10-174-579-183	Sequence 183, App
C 25	39.6	1.9	1594	9	US-10-174-582-183	Sequence 183, App
C 26	39.6	1.9	1594	9	US-10-174-588-183	Sequence 183, App
C 27	39.6	1.9	1594	9	US-10-175-739-183	Sequence 183, App
C 28	39.6	1.9	1594	9	US-10-175-740-183	Sequence 183, App
C 29	39.6	1.9	1594	9	US-10-175-743-183	Sequence 183, App
C 30	39.6	1.9	1594	9	US-10-176-488-183	Sequence 183, App
C 31	39.6	1.9	1594	9	US-10-176-492-183	Sequence 183, App
C 32	39.6	1.9	1594	9	US-10-176-747-183	Sequence 183, App
C 33	39.6	1.9	1594	9	US-10-176-750-183	Sequence 183, App
C 34	39.6	1.9	1594	9	US-10-176-985-183	Sequence 183, App
C 35	39.6	1.9	1594	9	US-10-176-987-183	Sequence 183, App
C 36	39.6	1.9	1594	9	US-10-176-991-183	Sequence 183, App
C 37	39.6	1.9	1594	9	US-10-176-992-183	Sequence 183, App
C 38	39.6	1.9	1594	9	US-10-176-993-183	Sequence 183, App
C 39	39.6	1.9	1594	9	US-10-184-658-183	Sequence 183, App
C 40	39.6	1.9	1594	9	US-10-173-695-183	Sequence 183, App
C 41	39.6	1.9	1594	9	US-10-173-697-183	Sequence 183, App
C 42	39.6	1.9	1594	9	US-10-173-705-183	Sequence 183, App
C 43	39.6	1.9	1594	9	US-10-174-576-183	Sequence 183, App
C 44	39.6	1.9	1594	9	US-10-174-585-183	Sequence 183, App
C 45	39.6	1.9	1594	9	US-10-174-586-183	Sequence 183, App

ALIGNMENTS

RESULT 1
US-09-864-761-11083
; Sequence 11083, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US 60/207,456
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661

NUMBER OF SEQ ID NOS: 49117
ANNOTATED SEQUENCE LISTING Engine vers. 1.1

; PRIOR FILING DATE: 2004 01 28

;
;
;

NUMBER OF SEQ ID NOS: 4911/
SOFTWARE: Annonax Sequence Listing Engine vers. 1.1

PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR APPLICATION NUMBER: 2001-01-30
; FILING DATE: 2001-01-30

ENCLOSURE


```

; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21956
; LENGTH: 256
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004468.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
; OTHER INFORMATION: SWISSPROT HIT: P11532, EVALUE 2.00e-30
; OTHER INFORMATION: NT HIT: gi5032282, EVALUE 1.00e-103
US-09-864-761-21956

```

```

Query Match
Best Local Similarity 6.9%; Score 140.2; DB 10; Length 256;
Matches 142; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAGAAGTATTATCGTGGCTTCTTTC 60
Db 51 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAGAAGTATTATCGTGGCTTCTTTC 60
QY 61 TGCTGAGGACACATTGCAAGCACAAAGGAGAGATTCTTAATGATGTGGAAGTGGTGAAGA 110
Db 111 TGCTGAGGACACATTGCAAGCACAAAGGAGAGATTCTTAATGATGTGGAAGTGGTGAAGA 120
QY 121 CCAGTTTCATCTCATGAGGGTAC 145
Db 171 CCAGTTTCATCTCATGAGGTAAC 195

```

```

RESULT 4
US-09-864-761-6092
Sequence 6092, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.

```

```

; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 6092
; LENGTH: 466
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004468.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
US-09-864-761-6092

```

```

Query Match
Best Local Similarity 6.1%; Score 124; DB 10; Length 466;
Matches 124; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAGAAGTATTATCGTGGCTTCTTTC 60
Db 343 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAGAAGTATTATCGTGGCTTCTTTC 60
QY 61 TGCTGAGGACACATTGCAAGCACAAAGGAGAGATTCTTAATGATGTGGAAGTGGTGAAGA 120
Db 403 TGCTGAGGACACATTGCAAGCACAAAGGAGAGATTCTTAATGATGTGGAAGTGGTGAAGA 462

```


y	121	CCAG	124
b	463	CCAG	466

RESULT 5
 US-09-960-352-10742/C
 ; Sequence 10742, Application US/09960352
 ; Patent No. US20020137139A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Warren, Wesley C.
 ; APPLICANT: Tao, Nengbing
 ; APPLICANT: Wyatt, John C.
 ; APPLICANT: Mathialagan, Nagappan
 ; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
 ; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
 ; FILE REFERENCE: 16511.006/37-21(10298)C
 ; CURRENT APPLICATION NUMBER: US/09/960,352
 ; CURRENT FILING DATE: 2001-09-24
 ; NUMBER OF SEQ ID NOS: 15112
 ; SEQ ID NO 10742
 ; LENGTH: 436
 ; TYPE: DNA
 ; ORGANISM: Bos taurus
 ; OTHER INFORMATION: Clone ID: 46-LIB3058-032-Q1-K1-D10
 ; US-09-960-352-10742

```

US-09-960-352 10742
      Query Match      2.1%;      Score 43;      DB 10;      Length 436;
      Best Local Similarity 48.6%;
      Pred. No. 0.017;
      0; Mismatches 125; Indels 0; Gaps 0;
      Matches 118; Conservative

```

[illegible]

QY	454	CCA	456
DB	191	GCA	189

RESULT 6
US-09-864-761-18355
; Sequence 18355, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03

```

PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 18355
LENGTH: 423
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC010133.1
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 3.1
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.8
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.6
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 3.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 2.2
US-09-864-761-18355

```

[illegible]

RESULT 7

US-09-960-352-4010/c
; Sequence 4010, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 4010
; LENGTH: 425
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 18-BOVMS1-007-Q1-E1-E5
US-09-960-352-4010

Query Match

Best Local Similarity 2.0%; Score 40.6; DB 10; Length 425;
Matches 133; Conservative 0; Mismatches 154; Indels 0; Gaps 0;
QY 220 AAAATTATCAGAAAGATGAAGAACTGAAGTACAAGAGCAGATGAATCTCTAAATTCAG 279
Db 424 AAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 279
QY 280 ATGGGAATGCCTCAGGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAG 365
Db 364 AAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 339
QY 340 GGATCTCCAGAAATCAGAAACTGAAGAGTTGAATGACTGGCTAAACAAAACAGAGAAAG 305
Db 304 AAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 399
QY 400 AACAAGGAAATGGAGGAAGAGGCTCTTGGACCTGATCTTGAAGACCTAAACAGCCCAAGT 245
Db 244 AACAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 459
QY 460 ACAACACACATAAGGTGCTTCAAGAAAGATCTAGAACAAAGAACAAAGTCA 185
Db 184 AAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 138

RESULT 8

US-09-864-864-204
; Sequence 204, Application US/09864864
; Patent No. US20020102679A1
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Secrist, Heather
; APPLICANT: Lodes, Michael J.
; APPLICANT: Algate, Paul A.
; APPLICANT: Fling, Steve P.
; APPLICANT: Mannion, Jane
; APPLICANT: Benson, Darin R.
; APPLICANT: Carter, Darrick
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.523
; CURRENT APPLICATION NUMBER: US/09/864,864
; CURRENT FILING DATE: 2001-05-23
; NUMBER OF SEQ ID NOS: 341
SOFTWARE: Corixa Invention Disclosure Database
SEQ ID NO 204

; LENGTH: 428

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-864-864-204

Query Match

Best Local Similarity 2.0%; Score 40.4; DB 10; Length 428;
Matches 86; Conservative 0; Mismatches 76; Indels 0; Gaps 0;
QY 258 AGATGAATCTCTAAATTCAGATGGGAATGCCCTCAGGGTAGCTAGCATGGAAAAACAAA 317
Db 208 ACAGGAAATTAAGAGAGCAACCATATAAGAGTCTCATGGGAGCAATGCATGAACAAACTG 267
QY 318 GCAATTTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACTGAAAGAGTTGAATGACT 377
Db 268 AGAATTTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACTGAAAGAGTTGAATGACT 377
QY 378 GGCTAACAAAAACAGAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 327
Db 328 GGACTCAAAAATTTAATGTGTCTCAACAGAGAAATCGATGAAG 369

RESULT 9

US-09-960-352-12395/c
; Sequence 12395, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 12395
; LENGTH: 418
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 53-LIB3058-009-Q1-K1-F2
US-09-960-352-12395

Query Match

Best Local Similarity 2.0%; Score 39.8; DB 10; Length 418;
Matches 122; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
QY 220 AAAATTATCAGAAAGATGAAGAACTGAAGTACAAGAGCAGATGAATCTCTAAATTCAG 279
Db 413 AAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 279
QY 280 ATGGGAATGCCTCAGGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAGCTAG 354
Db 353 ACAGATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 339
QY 340 GGATCTCCAGAAATCAGAAACTGAAGAGTTGAATGACTGGCTAAACAAAACAGAGAAAG 294
Db 293 ATAAAGAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 399
QY 400 AACAAGGAAATGGAGGAAGAGGCTCTTGGACCTGATCTTGAAGAGCTTAAACAGCCCAAGT 234
Db 233 AAAGAGAAAAATCCAATAATCGAAACACGACGCTCGTAATAAAGAGTCTCCAAAGCCTAGC 459
QY 460 ACAACACACATAAGGTGCTT 478
Db 173 ATCTAGACACAGAAATCCTT 155

RESULT 10

US-09-822-846-343/c
; Sequence 343, Application US/09822846
; Publication No. US20030027139A1

us-09-845-416-6_copy_1020_3060.rnpb

Mon Apr 28 09:28:36 2003

ORGANISM: Homo Sapien
US-10-174-590-183

GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
APPLICANT: McCoy, John M.
APPLICANT: Lavallie, Edward R.
APPLICANT: Collins-Racie, Lisa A.
APPLICANT: Evans, Cheryl
APPLICANT: Merberg, David
APPLICANT: Treacy, Maurice
APPLICANT: Agostino, Michael J.
APPLICANT: Steininger II, Robert J.
APPLICANT: Bowman, Michael R.
APPLICANT: Spaulding, Vikki
APPLICANT: Wong, Gordon G.
APPLICANT: Clark, Hilary
APPLICANT: Fechtel, Kim
APPLICANT: Howes, Steven H.
APPLICANT: Resnick, Richard J.
APPLICANT: Gulukota, Kamalakari
APPLICANT: Graham, James R.
APPLICANT: Genetics Institute, Inc.
TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
FILE REFERENCE: GIN 6400
CURRENT APPLICATION NUMBER: US/09/822,846
CURRENT FILING DATE: 2001-03-29
PRIOR APPLICATION NUMBER: 60/195,605
PRIOR FILING DATE: 2000-04-06
NUMBER OF SEQ ID NOS: 629
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 343
LENGTH: 1381
TYPE: DNA
ORGANISM: Xenopus sp.
US-09-822-846-343

Query Match 1.9%; Score 39.6; DB 9; Length 1381;
Best Local Similarity 60.0%; Pred. No. 0.39; Indels 0; Gaps 0;
Matches 66; Conservative 0; Mismatches 44;

QY 309 AAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAACTGAAAGAGT 368
1087 AAAACAAATTTAAATATTTTCAATTTTAAATTTGCTCCACCATAGATGAATAAAGAGC 1028
DB 1087 TGAATGACTGGCTAACAAACACAGAAAGAAAGAAAGAAATGGAGGAA 418
QY 369 TGAATGACTGGCTAACAAACACAGAAAGAAAGAAAGAAATGGAGGAA 418
DB 1027 TTAAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGAA 978

RESULT 11
US-10-174-590-183/c
Sequence 183, Application US/10174590
Publication No. US20030008352A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C42
CURRENT APPLICATION NUMBER: US/10/174,590
CURRENT FILING DATE: 2002-06-18
Prior application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 183
LENGTH: 1594
TYPE: DNA

Query Match 1.9%; Score 39.6; DB 9; Length 1594;
Best Local Similarity 60.0%; Pred. No. 0.43; Indels 0; Gaps 0;
Matches 66; Conservative 0; Mismatches 44;

QY 309 AAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAACTGAAAGAGT 368
1281 AAAACAAATTTAAATATTTTCAATTTTAAATTTGCTCCACCATAGATGAATAAAGAGC 1222
DB 1281 TGAATGACTGGCTAACAAACACAGAAAGAAAGAAAGAAATGGAGGAA 418
QY 369 TGAATGACTGGCTAACAAACACAGAAAGAAAGAAAGAAATGGAGGAA 418
DB 1221 TTAAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGATGGA 1172

RESULT 12
US-10-176-758-183/c
Sequence 183, Application US/10176758
Publication No. US20030008353A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
TITLE OF INVENTION: ACIDS ENCODING THE SAME
FILE REFERENCE: P3430R1C104
CURRENT APPLICATION NUMBER: US/10/176,758
CURRENT FILING DATE: 2002-06-21
Prior application removed - See File Wrapper or Palm
NUMBER OF SEQ ID NOS: 612
SEQ ID NO 183
LENGTH: 1594
TYPE: DNA
ORGANISM: Homo Sapien
US-10-176-758-183

Query Match 1.9%; Score 39.6; DB 9; Length 1594;
Best Local Similarity 60.0%; Pred. No. 0.43; Indels 0; Gaps 0;
Matches 66; Conservative 0; Mismatches 44;

QY 309 AAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAACTGAAAGAGT 368
1281 AAAACAAATTTAAATATTTTCAATTTTAAATTTGCTCCACCATAGATGAATAAAGAGC 1222
DB 1281 TGAATGACTGGCTAACAAACACAGAAAGAAAGAAAGAAATGGAGGAA 418
QY 369 TGAATGACTGGCTAACAAACACAGAAAGAAAGAAAGAAATGGAGGAA 418
DB 1221 TTAAGGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGGATGGA 1172

RESULT 13
US-10-175-737-183/c
Sequence 183, Application US/10175737
Publication No. US20030013153A1
GENERAL INFORMATION:
APPLICANT: Baker, Kevin P.
APPLICANT: Chen, Jian
APPLICANT: Desnoyers, Luc
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Gurney, Austin L.
APPLICANT: Pan, James
APPLICANT: Smith, Victoria
APPLICANT: Watanabe, Colin K.
APPLICANT: Wood, William I.

RESULT 15
US-10-175-738-183/c
: Sequence 183, Application US/10175738
: Publication No. US20030022294A1

```

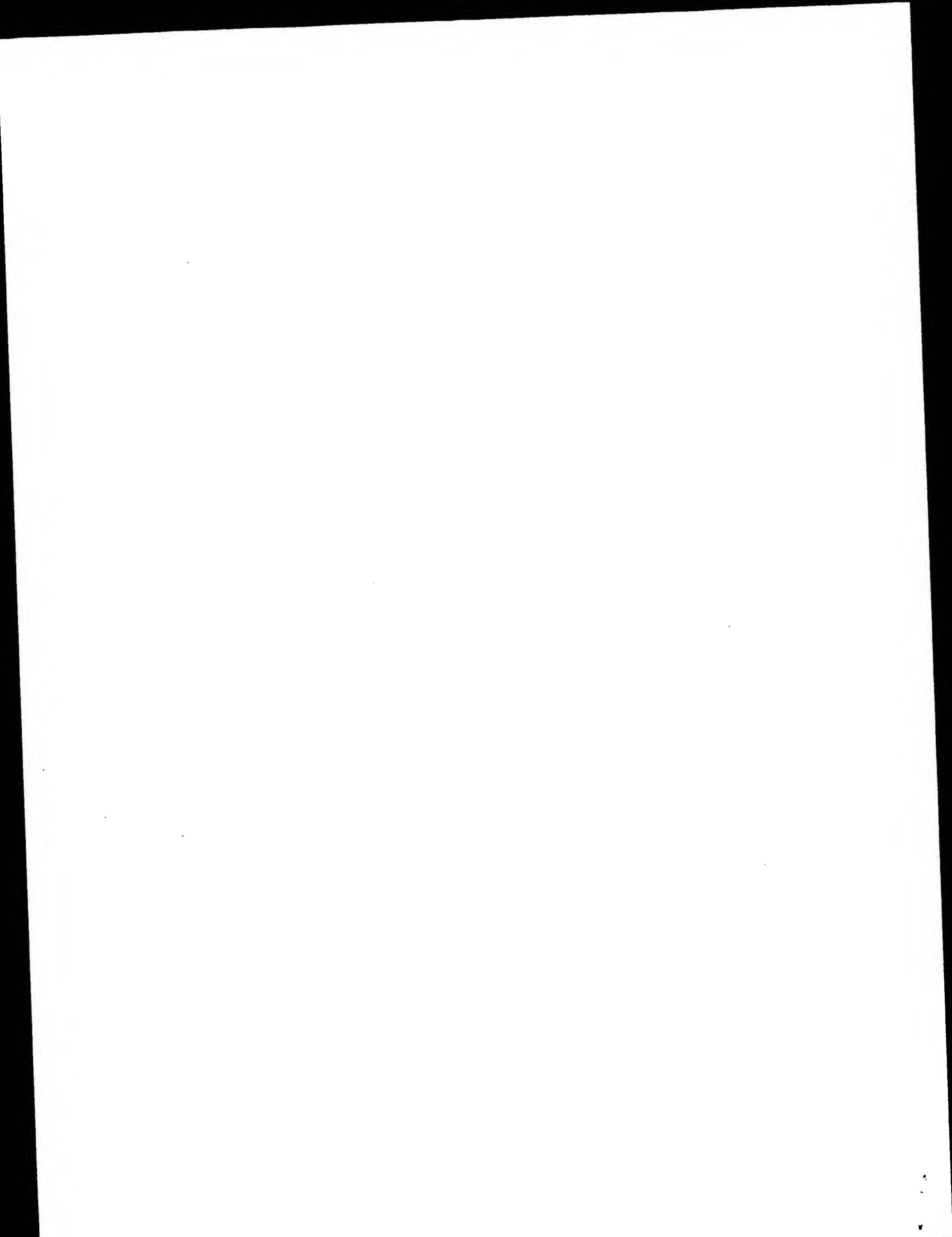
Query Match          1.9%;   Score 39.6;   DB 9;   Length 1594;
Best Local Similarity 60.0%;   Pred. NO. 0.43;
Matches 66;   Conservative 0;   Mismatches 44;   Indels 0;   Gaps 0;

QY  309  AAAACACAAAGCAATTTACATAGAGTTTTAAATGGATCTCCAGAAATCAGAAACTGAAACAGT 368
      |||||  ||  ||  ||  |||||  |||||  ||  |||  |||||
Db  1281  AAAACACAATTTAAATATTTCAAATTTTAAATTTCTCCACCATAAGATGAATAAAGAGC 1222
      |||||  |||||  |||||  |||||  |||||  |||||  |||||

QY  369  TGAATGACTGGCTAACAAAACAGAAAGAAAGAAACAAAGGAAAATGGAGGAA 418
      ||  ||  ||  ||  |||||  |||||  |||||  ||  ||  ||  ||  ||  ||
Db  1221  TTACTTAAAGGAAAGAAACAAAACAAACAAAAAACAACCGGATGGA 1172
      |||||  |||||  |||||  |||||  |||||  |||||  |||||

Search completed: April 25, 2003, 23:58:05
Job time : 137.776 secs

```

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 07:56:07 ; Search time 4707.23 seconds
(without alignments)
17070.103 Million cell updates/sec

Title: US-09-845-416-9_COPY_540_3300
Perfect score: 2761
Sequence: 1 gccagacatttgcactgga.....aagtggaagttcaacagga 2761

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 3813070

Minimum DB seq length: 0
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1281	46.4	4402	6	E30219	E30219 Shortened d
2	1168.6	42.3	4402	6	E30220	E30220 Shortened d
3	1078.8	39.1	4075	6	E30221	E30221 Shortened d
4	769.2	27.9	3747	6	E30218	E30218 Shortened d
5	713.4	25.8	3521	5	AF339031	AF339031 Danio rer
6	685	24.8	2654	5	FSCDYSTRO	M37645 Torpedo cal
7	644.4	23.3	3275	10	MUSDYS	M18025 Mouse dyster
8	604.8	21.9	3161	10	MMGUTRPH	X83506 M.musculus
9	565.6	20.5	3163	6	E30223	E30223 Shortened d
10	523.4	19.0	3073	10	AF195788	AF195788 Rattus no
11	523.4	19.0	3140	10	AF195787	AF195787 Rattus no
12	523.4	19.0	3499	9	HSU43519	U43519 Human dyster
13	476.4	17.3	630	9	HSDMDF1	X06179 Human fetal
14	457	16.6	2110	9	HUMDMXX	M92650 Human Duche
15	457	16.6	4658	9	BC028720	BC028720 Homo sapi
16	451	16.3	3172	6	E30222	E30222 Shortened d
17	421	15.2	1240	10	RNAPDY3	X69767 R.norvegicu
18	354.8	12.9	3650	3	SPAJ3356	AJ223356 Strongylo
19	307.6	11.1	1961	5	XLDYSTROP	X99700 X.laavis mR
20	286.8	10.4	1195	5	AF339032	AF339032 Danio rer
21	275.4	10.0	1993	5	SCDYSTROP	X99702 S.caniculua
22	273.4	9.9	500	9	HUMDYSTR15	L05649 Homo sapien
23	273.2	9.9	1898	10	BC024140	BC024140 Mus muscu
24	241.6	8.8	384	10	RNADYI	X65468 R.norvegicu
25	237.2	8.6	506	10	AB011666	AB011666 Rattus no
26	220.2	8.0	1966	9	HSDMDA1	X06178 Human adult
27	210.6	7.6	1428	5	SCU43517	U43517 Scylliorhinu
28	203	7.4	351	9	HUMDYSTO6	M86889 H.sapiens d
29	201.2	7.3	1482	5	XLUTROPHI	X99701 X.laavis mR
30	198.2	7.2	1500	6	AX107969	AX107969 Sequence
31	198.2	7.2	1500	6	AX107970	AX107970 Sequence
32	194.2	7.0	645	9	AF213410	AF213410 Homo sapi
33	192.6	7.0	1737	5	SCUTROPHI	X99703 S.caniculua
34	187	6.8	525	9	AF213411	AF213411 Homo sapi
35	186.4	6.8	306	9	AF213408	AF213408 Homo sapi
36	186.4	6.8	1249	9	HSDMD	X06293 Human Duche
37	179.8	6.5	374	10	S62620	S62620 (mdx3Cv)=dy
38	179	6.5	179	9	HUMDMDE	M23261 Human Duche
39	177.8	6.4	1440	5	GSU43518	U43518 Gobius sp.
40	175.8	6.4	360	9	HUMDYSTR13	L05647 Homo sapien
41	169.6	6.1	1428	10	MMU43520	U43520 Mus musculu
42	168	6.1	618	9	AF213440	AF213440 Homo sapi
43	160.4	5.8	343	9	HUMDYSDMD	L04186 Homo sapien
44	157	5.7	300	9	S38776	S38776 Homo sapien
45	154.6	5.6	1431	3	PSDYSTROP	X99738 Pectinidae

ALIGNMENTS

RESULT 1					
E30219					
LOCUS	E30219				
DEFINITION	Shortened dystrophin.				
ACCESSION	E30219				
VERSION	E30219.1 GI:13017026				
KEYWORDS	JP 1999318467-A/2.				
SOURCE	unidentified.				
ORGANISM	unidentified				
REFERENCE	1 (bases 1 to 4402)				
AUTHORS	Sinichi,T.				
TITLE	Shortened dystrophin				
JOURNAL	Patent: JP 1999318467-A 2 24-NOV-1999;				
	SCIENCE & TECH AGENCY,NATIONAL CENTER OF				
	NEUROLOGY AND PSYCHIATRY				

4402 bp	DNA	linear	PAT 18-JUN-2001
---------	-----	--------	-----------------

QY 1861 AAAATAGATGAGACCCCTTGAAGACTCCAGGAACTTCAAGAGGCCACGGATGAGCTGGA 1920
Db 1860 AAAATAGATGAGACCCCTTGAAGACTCCAGGAACTTCAAGAGGCCACGGATGAGCTGGA 1919
QY 1921 CCTAAGCTGCGCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGATCTCCT 1980
Db 1920 CCTAAGCTGCGCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGATCTCCT 1979
QY 1981 CATTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAATTCGCC 2040
Db 1980 CATTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAATTCGCC 2039
QY 2041 TCTGAAAGAGAACGCTGAGCCACGCTCAATGACCTTGCTCGCCAGCTTACCACCTTTGGGCAT 2100
Db 2040 TCTGAAAGAGAACGCTGAGCCACGCTCAATGACCTTGCTCGCCAGCTTACCACCTTTGGGCAT 2099
QY 2101 TCAGCTCTCACCGTATACCTCAGCACTCTGGAAGACCTGAACACCAAGATGGAAGCTTCT 2160
Db 2100 TCAGCTCTCACCGTATACCTCAGCACTCTGGAAGACCTGAACACCAAGATGGAAGCTTCT 2159
QY 2161 GCAGGTGCGCTCGAGGACCGAGTCAGGCAGCTGCATGAAGCCACAGGGACTTTGGTCC 2220
Db 2160 GCAGGTGCGCTCGAGGACCGAGTCAGGCAGCTGCATGAAGCCACAGGGACTTTGGTCC 2219
QY 2221 AGCATCTCAGCACTTTCTTCCACGCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCC 2280
Db 2220 AGCATCTCAGCACTTTCTTCCACGCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCC 2279
QY 2281 AAACAAAGTGCCCTACTATATCAACCCAGGACTCAACAAACTTGCTGGGACCATCCCAA 2340
Db 2280 AAACAAAGTGCCCTACTATATCAACCCAGGACTCAACAAACTTGCTGGGACCATCCCAA 2339
QY 2341 AATGACAGAGCTCTACCACTCTTTAGCTGACCTGAATAATGTGAGATTTCTCAGCTTATAG 2400
Db 2340 AATGACAGAGCTCTACCACTCTTTAGCTGACCTGAATAATGTGAGATTTCTCAGCTTATAG 2399
QY 2401 GACTGCCATGAACCTCCGAAGACTGCAGAGGCTTGTCTGGATCTCTTGAGCTGTC 2460
Db 2400 GACTGCCATGAACCTCCGAAGACTGCAGAGGCTTGTCTGGATCTCTTGAGCTGTC 2459
QY 2461 AGCTGCATGTGATGCCTTGACAGCAACCTCAAGCAAAATGACCCAGCCCATGGATAT 2520
Db 2460 AGCTGCATGTGATGCCTTGACAGCAACCTCAAGCAAAATGACCCAGCCCATGGATAT 2519
QY 2521 CCTGCAGATTATTAATGTTTGAACCACTATTATGACCGCTGGAGCAAGAGCAACAA 2580
Db 2520 CCTGCAGATTATTAATGTTTGAACCACTATTATGACCGCTGGAGCAAGAGCAACAA 2579
QY 2581 TTTGGTCAACGCTCCCTCTCTGCGTGGATATGTGCTGAACTGGCTGCTGAATGTTTATGA 2640
Db 2580 TTTGGTCAACGCTCCCTCTCTGCGTGGATATGTGCTGAACTGGCTGCTGAATGTTTATGA 2639
QY 2641 TACGGACGAACAGGAGGATCCGTGCTCTTTTAAACTGGCATCATTTCCCTGTG 2700
Db 2640 TACGGACGAACAGGAGGATCCGTGCTCTTTTAAACTGGCATCATTTCCCTGTG 2699
QY 2701 TAAAGCACATTTGGAAGACAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGG 2760
Db 2700 TAAAGCACATTTGGAAGACAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGG 2759
QY 2761 A 2761
b 2760 A 2760

ESULT 2
30220
DCUS E30220
DEFINITION Shortened dystrophin.
CCESION E30220
E30220.1 GI:13017027
EYWORDS JP 1999318467-A/3.

4402 bp DNA linear PAT 18-JUN-2001

SOURCE ORGANISM unidentified.
unidentified
unclassified.
REFERENCE 1 (bases 1 to 4402)
AUTHORS Sinichi,T.
TITLE Shortened dystrophin
JOURNAL Patent: JP 1999318467-A 3 24-NOV-1999;
SCIENCE & TECH AGENCY,NATIONAL CENTER OF NEUROLOGY AND PSYCHIATRY
COMMENT OS Unidentified
PN JP 1999318467-A/3
PD 24-NOV-1999
PF 08-MAY-1998 JP 1998142134
PR SINICHI TAKEDA
PI C12N15/09,A61K48/00,C12N15/00
PC Strandedness: Both;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..4402
FT /organism='Unidentified'.
FEATURES
source Location/Qualifiers
1..4402
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 1339 a 984 c 1010 g 1069 t
ORIGIN
Query Match 42.3%; Score 1168.6; DB 6; Length 4402;
Best Local Similarity 71.1%; Pred. No. 3.6e-291;
Matches 1962; Conservative 0; Mismatches 64; Indels 735; Gaps 2;
QY 1 GCCAGACCTATTGACTGGAAATAGTGTGGTTGGCCAGCAGTCAGCCACACACGACTGGA 60
Db 735 GCCAGACCTATTGACTGGAAATAGTGTGGTTGGCCAGCAGTCAGCCACACACGACTGGA 794
QY 61 ACATGCATTCAACATCGCCAGATATCAATTAGGCATAGAGAAACTACTCGATCCTGAAGA 120
Db 795 ACATGCATTCAACATCGCCAGATATCAATTAGGCATAGAGAAACTACTCGATCCTGAAGA 854
QY 121 TGTGTATACCACTATCCAGATAAGAAAGTCCATCTTAATGTACATCACATCCTCTCCA 180
Db 855 TGTGTATACCACTATCCAGATAAGAAAGTCCATCTTAATGTACATCACATCCTCTCCA 914
QY 181 AGTTTGCCTCAACAAGTGAGCATTTGAAGCCATCCAGGAAGTGGAAATGTTGCCAAGGCC 240
Db 915 AGTTTGCCTCAACAAGTGAGCATTTGAAGCCATCCAGGAAGTGGAAATGTTGCCAAGGCC 974
QY 241 ACCTAAGTGACTAAAGAAACATTTTCAGTTACATCATCAATGCACATATTCTCAACA 300
Db 975 ACCTAAGTGACTAAAGAAACATTTTCAGTTACATCATCAATGCACATATTCTCAACA 1034
QY 301 GATCAGGTCAGTCTAGCACAGGGATATGAGAGAACTTCTTCCCTTAAGCCTCGATTCAA 360
Db 1035 GATCAGGTCAGTCTAGCACAGGGATATGAGAGAACTTCTTCCCTTAAGCCTCGATTCAA 1094
QY 361 GAGCTATGCCTACACACAGGCTGCTTATGTCAACCACCTCTGACCCCTACACGGAGCCCAT 420
Db 1095 GAGCTATGCCTACACACAGGCTGCTTATGTCAACCACCTCTGACCCCTACACGGAGCCCAT 1154
QY 421 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAGTCAATTTGGCAGTTTATTGATGGAGAG 480
Db 1155 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAGTCAATTTGGCAGTTTATTGATGGAGAG 1214
QY 481 TGAAGTAAACCTGGACCGTTATCAACACAGCTTTAGAAGAAGTATTATCGTGGCTTCTTC 540
Db 1215 TGAAGTAAACCTGGACCGTTATCAACACAGCTTTAGAAGAAGTATTATCGTGGCTTCTTC 1274
QY 541 TGCTGAGGACACATTTGCAAGCACACAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 600
Db 1275 TGCTGAGGACACATTTGCAAGCACACAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 1334
QY 601 CCAGTTTCACTACTCATGAGGGGTACATGATGGATTGACAGCCCATCAGGGCCGGTTGG 660
|||||


```
RESULT 3
E30221
LOCUS Shortened dystrophin. 4075 bp DNA linear PAT 18-JUN-2001
DEFINITION E30221
ACCESSION E30221
VERSION E30221.1 GI:13017028
KEYWORDS JP 1999318467-A/4.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 4075)
AUTHORS Sinichi,T.
TITLE Shortened dystrophin
JOURNAL Patent: JP 1999318467-A 4 24-NOV-1999;
SCIENCE & TECH AGENCY,NATIONAL CENTER OF NEUROLOGY AND PSYCHIATRY
COMMENT OS Unidentified
PN JP 1999318467-A/4
PD 24-NOV-1999
PF 08-MAY-1998 JP 1998142134
PR SINICHI TAKEDA
PI CL12N15/09,A61K48/00,C12N15/00
PC Strandedness: Both;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..4075
FEATURES
source Location/Qualifiers
1..4075 /organism='Unidentified'.
/organism='unidentified'
/db_xref="taxon:32644"
BASE COUNT 1237 a 910 c 936 g 992 t
ORIGIN
Query Match 39.1%; Score 1078.8; DB 6; Length 4075;
Best Local Similarity 99.4%; Pred. No. 6.3e-268;
Matches 1083; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 1 GCCAGACCTATTGACTGGATAGTGGTTGCCAGCAGTCCAGCCACACACGACTGGA 60
Db 735 GCCAGACCTATTGACTGGATAGTGGTTGCCAGCAGTCCAGCCACACACGACTGGA 794
QY 61 ACATGCATTCAACATCGCCAGATATCAATAGGCATAGAGAACTACTCGATCCTGAAGA 120
Db 795 ACATGCATTCAACATCGCCAGATATCAATAGGCATAGAGAACTACTCGATCCTGAAGA 854
QY 121 TGTGATACCACTATCCAGATAAAGTCCATCTTAATGTACATCATCATCTTTCCA 180
Db 855 TGTGATACCACTATCCAGATAAAGTCCATCTTAATGTACATCATCATCTTTCCA 914
QY 181 AGTTTGCCTCAACAAGTGAGCATTTGAAGCCATCCAGAAAGTGGAAATGTTGCCAAGGCC 240
Db 915 AGTTTGCCTCAACAAGTGAGCATTTGAAGCCATCCAGAAAGTGGAAATGTTGCCAAGGCC 974
QY 241 ACCTAAAGTGACTAAAGAAGAACATTTTCAGTTACATCATCAATGCACATTTCTCAACA 300
Db 975 ACCTAAAGTGACTAAAGAAGAACATTTTCAGTTACATCATCAATGCACATTTCTCAACA 1034
QY 301 GATCAGGTCAGTCTAGCACAGGGATATGAGAGAACTTCTCCCTAAGCCCTGATTCAA 360
Db 1035 GATCAGGTCAGTCTAGCACAGGGATATGAGAGAACTTCTCCCTAAGCCCTGATTCAA 1094
QY 361 GAGCTATGCCTACACACAGGCTGCTTATGTACACACCTCTGACCCCTACAGGAGCCCAT 420
Db 1095 GAGCTATGCCTACACACAGGCTGCTTATGTACACACCTCTGACCCCTACAGGAGCCCAT 1154
QY 421 TCCTTACAGCATTTGGAAGCTCCTGAAGACAAAGTCAATTTGGCAGTTCATTTGAGAG 480
Db 1155 TCCTTACAGCATTTGGAAGCTCCTGAAGACAAAGTCAATTTGGCAGTTCATTTGAGAG 1214
QY 481 TGAAGTAAACCTGGACCGTTATCAACAGCTTTAGAGAAGTATTCGTGGCTTCTTC 540
Db 1155 TGAAGTAAACCTGGACCGTTATCAACAGCTTTAGAGAAGTATTCGTGGCTTCTTC 540
```

```
Db 1215 TGAAGTAAACCTGGACCGTTATCAACAGCTTTAGAGAAGTATTCGTGGCTTCTTC 1274
QY 541 TGCTGAGGACACATTTGCAAGCACAAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 600
Db 1275 TGCTGAGGACACATTTGCAAGCACAAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 1334
QY 601 CCAGTTTCACTACTGAGGGGTACATGATGGATTGACAGCCCATCAGGGCCGGGTGG 660
Db 1335 CCAGTTTCACTACTGAGGGGTACATGATGGATTGACAGCCCATCAGGGCCGGGTGG 1394
QY 661 TAAATTTCTACAATTTGGGAAGTAAGCTGATTTGGAACAGGAAATTTATCAGAAGATGAAGA 720
Db 1395 TAAATTTCTACAATTTGGGAAGTAAGCTGATTTGGAACAGGAAATTTATCAGAAGATGAAGA 1454
QY 721 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGATGGGAATGCCCTCAGGGTAGC 780
Db 1455 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGATGGGAATGCCCTCAGGGTAGC 1514
QY 781 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAACTCAGAAACT 840
Db 1515 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAACTCAGAAACT 1574
QY 841 GAAAGAGTTGAATGACTGGCTAAACAAAACAGAAAGAAAGAAACAAAGAAATGGAGGAAGA 900
Db 1575 GAAAGAGTTGAATGACTGGCTAAACAAAACAGAAAGAAAGAAACAAAGAAATGGAGGAAGA 1634
QY 901 GCCTCTTGGACCTGATCTTGAAGACCTAAACGCCCAAGTACAAACATAAGGTGCTTCA 960
Db 1635 GCCTCTTGGACCTGATCTTGAAGACCTAAACGCCCAAGTACAAACATAAGGTGCTTCA 1694
QY 961 AGAAGATCTAGAACAAAGAACAAAGTCAAGGTCAATCTCTCACTCACATGGTGGTAGT 1020
Db 1695 AGAAGATCTAGAACAAAGAACAAAGTCAAGGTCAATCTCTCACTCACATGGTGGTAGT 1754
QY 1021 TGATGAATCTAGTGGAGATCAGCGAAGTCAAGGTCAATCTCTCACTCACATGGTGGTAGT 1080
Db 1755 TGATGAATCTAGTGGAGATCAGCGAAGTCAAGGTCAATCTCTCACTCACATGGTGGTAGT 1814
QY 1081 AGATCGATGG 1090
Db 1815 CACCAGATGG 1824
```

```
RESULT 4
E30218
LOCUS Shortened dystrophin. 3747 bp DNA linear PAT 18-JUN-2001
DEFINITION E30218
ACCESSION E30218
VERSION E30218.1 GI:13017025
KEYWORDS JP 1999318467-A/1.
SOURCE unidentified.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 3747)
AUTHORS Sinichi,T.
TITLE Shortened dystrophin
JOURNAL Patent: JP 1999318467-A 1 24-NOV-1999;
SCIENCE & TECH AGENCY,NATIONAL CENTER OF NEUROLOGY AND PSYCHIATRY
COMMENT OS Unidentified
PN JP 1999318467-A/1
PD 24-NOV-1999
PF 08-MAY-1998 JP 1998142134
PR SINICHI TAKEDA
PI CL12N15/09,A61K48/00,C12N15/00
PC Strandedness: Both;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..3747
FEATURES
source Location/Qualifiers
1..3747 /organism='Unidentified'.
/organism='unidentified'
```


SOURCE		Danio rerio.	
ORGANISM		Danio rerio	
		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.	
REFERENCE		1 (bases 1 to 3521)	
AUTHORS		Bolanos-Jimenez,F., Bordais,A., Behra,M., Strahle,U., Sahel,J. and Rendon,A.	
TITLE		Dystrophin and Dp71, two products of the DMD gene, show a different pattern of expression during embryonic development in zebrafish	
JOURNAL		Mech. Dev. 102 (1-2), 239-241 (2001)	
MEDLINE		21184125	
PUBMED		11287201	
REFERENCE		2 (bases 1 to 3521)	
AUTHORS		Bolanos-Jimenez,F., Rendon,A. and Strahle,U.	
TITLE		Direct Submission	
JOURNAL		Submitted (22-JAN-2001) Laboratoire de Physiopathologie Retinienne, EMI 99-18 INSERM-Universite Louis Pasteur, 1, Place de l'Hopital, Strasbourg 67091, France	
FEATURES		Location/Qualifiers	
source		1. .3521	
		/organism="Danio rerio"	
		/db_xref="taxon:7955"	
		/chromosome="1"	
		/map="between z5508 and z5058"	
		<1. .3521	
gene		/gene="dmd"	
		<1. .3292	
		/gene="dmd"	
		/note="alternatively spliced"	
		/codon_start=2	
		/product="dystrophin"	
		/protein_id="AAK38376.1"	
		/db_xref="GI:13699250"	
		/translation="RFCRHWEDSHAKLTARVLTQLNMYKSDSDWLEARKRVEPLIKKA NEKLESWKVSHVEDLKGQNAVVKQISKDLQWQTMNVNTELANKLLTLYADDDTS KVKQMTESMNLAWANIKKRAGDEADLEAGLRQLQHYLLDLEKFLNWLTAETANVL QDATEGLENPATVRLLEOWDLOAEIDAHRETYHSLDENGHRIVSSLEGTNAV VLQKRLDDMGQRWHELCKNKMVSIRPYLDAGVDQWKHLHMSLQELLNLWLQKRELEKQ KPVGGDPTVHQQLLTHKAFRELGAKEPVINGTLDNAKTFLAEMPREGLKQRPQKD VSPEERVQNVGRILRKEVEDVTVRWKNLGAASVDWQQQLELALERLMELODAQOLDY KLRAESVKNWKPVGELLVDDLQNHIDRVKAFQEEIAPIQDNVNVHVNQLASTFRPSD IQLSPDNLSRIDDLNMRWRLQISIEHLSQLTAFKDLGPSQNFHSAVESPLERSI SPNNVPYINHQTQTTCDWHPKMAELYQSLADLNVRFSAYRTAMKLRMRQKALCLDL LSPAAACEAFQHNKQNEQFMDIVQVINCITSYDRLEQQHSSLVNPLCVDMLNW LLNVYDTGRAGKIRTLSEKTIISLCKAHLEDKYRFLFRVASATGFCQDQRRLLGLLH DAIQIPRQLGEVASFSGSNIEPSVRSFCFANNKPELEASVFLDMRLEPQSMVWLPV LHRVAAAEATAKHAQCNICEKPIIGFRYRSKLFHFNVDICQSCFFSGRVAKGHKMOPY MVEYCTPTTSGEDVDFEAKVLKNKFKRYFAKHPRMGYLPVQTILEGDNMETPVTLI NFVPVDHPPASSPOLSHDDTHSRIEHYASRDDHLLLIQHYCQSLNQGSPLSQPSPAQ ILISMETEEKGELEVLNDLEQENRKLQAEYDRLKAHDKGLSPLSPQMLPVSPQ SPRDAELIAEAKLLRQHKRLEARMQILEDHKNQLESQTLRLROLLEQTESKVNQGTAL SSPSTASPRSDTSLASLRVAASQTTETMGDDDELSSPTQDASTGLEDVIEQLNNSFPHS QGGGRLLNP"	
3'UTR		3293. .3521	
		/gene="dmd"	
BASE COUNT		1005 a 868 c 892 g 756 t	
ORIGIN			
Query Match		25.8%; Score 713.4; DB 5; Length 3521;	
Best Local Similarity		62.5%; Pred. No. 2.4e-173;	
Matches 1151; Conservative		0; Mismatches 686; Indels 6; Gaps 2;	
QY	918	TTGAAGACCTAAACGCCAAGTACAAACATAGGTGCTTCAAGAAGATCTAGAACAAAG	977
Db	174	TTGAAGATTTAAAGAGGCCAAAATGCAGATGTCAAGCAACTTCTTAAGGATCTACAGCAGT	233
QY	978	AACAAGTCAGGGTCAATCTCTCACTCACATGGTGGTGTAGTGTAGTGAATCTAGTGGAG	1037
Db	234	GGCAGACTCAGATGAATGTCATAATGAGCTGGCAACAACTACTGACTCTTTATGCAG	293
QY	1038	ATCACGCAACTGCTGCTTTTGGAAAGAACAACTTAAGGTATTGGAGATCGATGGCAACA	1097

Db 294 ACACGACACAAAGCAAGTGAACAAATGACCCGAGAGCATGAATCTGGCTTGGGCCAAC 353
QY 1098 TCTGTAGATGGACAGAACCGCTGGGTCTCTTTTACAAGACACTCATAGATTACTGCAAC 1157
Db 354 TCAAAAAGCGTGCAGGGGACAAGGAGGAGATCTGGAAGCTGGACTTCGGCAGTTGCAGC 413
QY 1158 AGTTCCCTGGACCTGGAAAAGTTTCTTCTGCTGGCTTACAGAAGCTGAAACAACTGCCA 1217
Db 414 ATTACTACTGGACCTGGAGAAATTCCTTAATTTGGCTAACGGAGGAGCAAGAACACAGCCA 473
QY 1218 ATGTCTTACAGGATGCTACCCGTAAGGAAAGGCTCCTAGAAAGACTCCAAAGGAGTAAAG 1277
Db 474 ATGTCTTACAAGATGCCACCTTTAAGGAGGACTTCTGGAAATCCTGCCACAGTTGCAG 533
QY 1278 AGCTGATGAACAATGGCAAGACTCCAAAGGTGAAATTTGAAGCTCACACAGATGTTATC 1337
Db 534 ATTTACTCGAGCAATGGCAGGATCTCCAGGCAGAGATTGATGCTACCCGGGAGACGTACC 593
QY 1338 ACAACCTGGATGAAAACAGCAAAAAATCCTGAGATCCCTGGAAGGTTCGATGATGCAG 1397
Db 594 ATTCATTAGATGAAAACGACATCGTATTGTGCTCTCTGGAGGGAACGGACAATGCTG 653
QY 1398 TCCTGTTACAAAGACGTTTGGATAACATGAACCTTCAAGTGGAGTGAACCTCGGAAAAAGT 1457
Db 654 TGGTGCTACAAAACCGCTTGATGACATGGGGCAGCGCTGGCATGAGTTGTGCAACAAAG 713
QY 1458 CTCTCAACATTAGTCCCATTTTGGAGGCCAGTTCTGACCAGTGGAAAGCTTGCACCTTT 1517
Db 714 TTATGAGTATAAGGCCCTATCTAGATGCCCGGCTTGATCAGTGGAAACACTTGCACATGT 773
QY 1518 CTCTGCAGAACTTCTGGTGTGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCAC 1577
Db 774 CCTTACAAGAGCTGCTCAACTGGCTGCAGCTGAAGAGGGAAGAGCTGGAGAAGCAGAAGC 833
QY 1578 CTATTGGAGGCGACTTTCAGCAGTTCAGAAGCAGAACGATGTACATAGGGCCTTCAAGA 1637
Db 834 CAGTAGGGGCGACGTCGCCGACCGTTCACCAACAACCTCCTCAGCACAAGGCTTTCAGAA 893
QY 1638 GGAATTTGAAAACCTAAGAACCTGTAAATCATGAGTACTCTTGAGACTGTACGAATATTC 1697
Db 894 GAGAGCTGGGTGCCAAAGAACCTGTAAATCAATGGAACCTCTCGACAATGCGAAAACCTTC 953
QY 1698 TGACAGAGCAGCCTTTGGAAGGACTAGAGAAACTCTACCAGGAGCCAGAGAGCTGCCTC 1757
Db 954 TTGCTGAGATGCCCTCGTAGGGCCT---GAAACAGAGAGCTGGACAAAAGATGTAAGTC 1010
QY 1758 CTGAGGAGAGAGCCAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGTCAATA 1817
Db 1011 CTGAGGAGCGGTTCAAAAACGTTGGCCGCGATATTGCGTAAAGAAGTGGAGGATGTAACG 1070
QY 1818 CTGAGTGGGAAAATTTGAACCTGCACCTCCGCTGACTGGCAGAGAAAATAGATGAGACCC 1877
Db 1071 TGGATGGAAGAATCTGGGTGCAGCTTCTGTGAGCTGGCAGCAACAGCTGGAACTGGCCT 1130
QY 1878 TTGAAGAGCTCCAGGAACCTTCAAGAGGCGCACCGGATGAGCTGGACCTCAAGCTGCGCCAAG 1937
Db 1131 TGGAGAGGCTGATGGAGCTCCAGGACGCCAGGATCAGCTGGACTACAAGCTACGACAGG 1190
QY 1938 CTGAGGTGATCAAGGATCCTGGCAGCCCGTGGGCGATCTCCCTCAATGACTCTCTCCAAG 1997
Db 1191 CTGAGAGCGTGAAGAATCTTTGGAAGCCTGTTGGGAGCTGCTGTAGACGATTTGCAAA 1250
QY 1998 ATCACTCGAGAAAGTCAAGGACTTCGAGGAGAAATTTGGCGCTCTGAAAGAGAACGTGA 2057
Db 1251 ACCACATAGACAGAGTGAAGGCAATTCAGGAAGAGATAGCTCCCATTCAGGATAACGTAA 1310
QY 2058 GCCACGTAATGACCTTGTCTGCCAGCTTACCACCTTTGGGCAATTCAGCTCTCACCGTATA 2117
Db 1311 ATCATGTTAACCAAGCTGGCCTCCACATTTAGACCCGTGTGACATTCAGCTTCTCCAGACA 1370
QY 2118 ACCTCAGCAGCTTGGAAAGACCTGAACACACAGATGGAAGCTTCTGCAGGTGGCCGTCGAGG 2177
Db 1371 ACCTGAGCAGATCGATGACCTCAACATGAGATGGAGGCTTCTGCAGATCTCCATGAGG 1430

QY 2178 ACCGAGTCAGGCAGCTGCATGAAGCCCACAGGGACTTTGGTCCAGCATCTCAGCACTTTC 2237
Db 1431 AGCATCTGAGTCAGCTGACCACAGCCTTTAAGGACTTGG---ACCTCTCAGAATTTCC 1487
QY 2238 TTTCCACGCTCTGCCAGGTCCTCTGGGAGAGAGSCCATCTCGCCAAACAAAGTGCCCTACT 2297
Db 1488 TTCATGGCTCTGTTGAAAGCCCTCTCGAACGCTCTATTTTACCCCAACAATGTCCCTACT 1547
QY 2298 ATATCAACACAGAGACTCAAAACAACCTTCTGGGACCATCCCCAAATGACAGAGCTTACC 2357
Db 1548 ATATCAATCACAGACCCCAACAACATGTTGGACCCACCCAAAGATGGCAGAACTTACC 1607
QY 2358 AGTCTTTAGCTGACCTGAATAATGTGAGATTTCTCAGCTTATAGGACTGCCATGAAACTCC 2417
Db 1608 AGTCATTAGCGGATCTCAACAACGCTGCGGTTCTCGGCATACAGGACGCAATGAAGCTCA 1667
QY 2418 GAAGACTGCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCCTGTGAGCTGATGATGCCT 2477
Db 1668 GACGAATGCAGAAAGCCCTCTGTTTGGATCTTCTGAGCATGCCCTGAGCCCTGTGAAGCCT 1727
QY 2478 TGGACAGCACAACCTCAAGCAAAATGACCAGCCCATGGATATCTTCAACGCTCCCTC 2537
Db 1728 TTGAGCAGCACAATCTCAAAACAGAACGAGCAGTTTCATGGACATCGTGCAGGTGATCACT 1787
QY 2538 GTTTGACCACCTATTTATGACCCGCTTGGAGCAAGAGCAGCAGACAGCAGCTGGTCAACG 1847
Db 1788 GTCTGACCCAGCATCTACGACCGCTTGGAGCAGCAGCAGCAGCAGCTGGTCAACGCTGCCTC 1847
QY 2598 TCTGCTGGATGATGTCTGAACTGGCTGGCTGCTGATGATGTTTATGATACGGGACGACAGCA 2657
Db 1848 TCTGTGTGACATGTGTCTCAACTGGCTGCTCAACGTTTACGATACAGGACGAGCTGGA 1907
QY 2658 GGATCCGTGCTCTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCAGCATTTGGAAG 2717
Db 1908 AGATTGCTACCTTATCTTCAAAACAGAAATAATCTCTTTGTGCAAGCTCACCTTGAAG 1967
QY 2718 ACAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGG 2760
Db 1968 ATAAGTACAGATTTTATTTCGAGAGGTGGCCAGTGCCACAGG 2010

RESULT 6
FSCDYSTRO

LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

FSCDYSTRO
Torpedo californica dystrophin mRNA, 3' end.
M37645
M37645.1 GI:397970
dystrophin.
Torpedo californica,
Torpedo californica
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
Elasmobranchii; Squalea; Hypnosqualea; Pristiorajae; Batoidae;
Torpediniformes; Torpedinoidei; Torpedinidae; Torpedo.

REFERENCE
AUTHORS

TITLE
JOURNAL
MEDLINE
PUBMED

REFERENCE
AUTHORS
TITLE

JOURNAL
MEDLINE
PUBMED

COMMENT

1 (bases 1 to 2654)
Yeadon,J.E., Lin,H., Dyer,S.M. and Burden,S.J.
Dystrophin is a component of the subsynaptic membrane
J. Cell Biol. 115 (4), 1069-1076 (1991)
92064638
1720119
2 (bases 1 to 2654)
Ravin,A.J., Dyer,S.M., Yeadon,J.E. and Burden,S.J.
Multiple dystrophin isoforms are associated with the postsynaptic
membrane of Torpedo electric organ
J. Physiol. (Paris) 85 (3), 131-133 (1991)
92291902
1818109
On Sep 8, 1993 this sequence version replaced gi:213232.
Draft entry and computer-readable sequence for [Unpublished (1990)]
by S.J.Burden, 06-AUG-1990.
Biology Dept, 16-820
MIT

FEATURES
source
gene
CDS
Cambridge, MA 02139.
Location/Qualifiers
1. .2654
/organism="Torpedo californica"
/db_xref="taxon:7787"
1. .2654
/gene="dystrophin"
<1. .2654
/gene="dystrophin"
/codon_start=3
/product="dystrophin"
/protein_id="AAC38002.1"
/db_xref="GI:397971"
/translation="SGEQWKRLLQISLQDFLTWNLKNDLRRQPIGGDAPTVCQOND
VHRIFKRELKAKEPVMSALDVTVHLFLADPAIRAPESLLTGPREKIPENIQNVAKRI
RKYAEVKVENDKLSNRVSDWKRIDEALKRLLELQSDMDELNLKLRQAEAIKDTWOP
VGLLLDSLDQDHIKRVKFAEIAIPMKENVTMNDLASQFTPPDIQLSPYNLQLEDL
NTRWKLQVSIIDELLKQLHEAHRDFGPTSQHFLSTVQGPWEAISPKNKPYIINHQT
QTTCDWHPKMTLEYQSLADLNNVRFSAYRTAMKLRRLKQKALCLDLLSLPSACEAFDQH
NLKQNDQLLDILEIINCLTSIYDRLEQEHSLNVVPLCLVDMCLNLLNLYDTGRTGKI
RVLSFKSGIMSCKAHLEDKYRFLFKQVASPTGFCQRRRLGLLLHEAIIQIPRLGEVA
SPFGSNIETSVRSQFANNKPEIAALFLDWMRLEPQSLVWMPVLRVAAAETAKHQ
AKNICKEPIIGFRYSLKHFNYDVQSCFFSGRTAKGKMHYPMVEYCTPTTSGED
VRPAKLVKNFKRYFAKHPRMGYLPVQTVLEGDNLETPTVLIWFVPDYEPASSP
QLSHDDTHSRIEHLRSRLAEMNRNGSYLNDISIPNESIDDEHLLIHYCQSLNQESP
LSQPSPAQILISLESEERGELEERILADLEENRLQSEYEKLLQOHDHKGSLPLSP
PEMPPISQSPRDAELIAEAKLLRQHKGRLEARMQILEHNLKQLESQHLRLRQLLEQ
QAEVRNGTSVSSPSTSSQSDSQPVLHGVGSQTSGLGEDDLLSAPQSTSELED
VMEQLSSFPSSQRRDTPGKKMKVPM"

BASE COUNT 773 a 618 c 646 g 617 t
ORIGIN
Query Match 24.8%; Score 685; DB 5; Length 2654;
Best Local Similarity 71.0%; Pred. No. 5.2e-166;
Matches 907; Conservative 0; Mismatches 370; Indels 0; Gaps 0;
QY 1485 CCAGTCTGTGACCAAGTGGAGAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGTGGCTAC 1544
Db 1 CAAGTGGTGAGCAGTGGAGAGCGGCTGCAGATCTCCCTCCAGGACTTCTGTGACGTGGATGA 60
QY 1545 AGCTGAAGATGATGAATTAAGCCGCGAGGACCTATTGGAGGCGACTTTCCAGCAGTTC 1604
Db 61 ATTTGAAGAACGATGAACCTAAGGAGGCAATGCCATTTGGTGGTGATGCTCCACAGTCT 120
QY 1605 AGAAGCAGAACGATGTACATAGGCGCTTCAAGAGGGAATTTGAAACCTAAGAAACCTGTAA 1664
Db 121 GTCAGCAGAACGATGTTCACCGAATTTTCAAGCAGAGATTTGAAGGCAAGAAACAGTGG 180
QY 1665 TCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTTGAAGGACTAG 1724
Db 181 TCATGAGCGCTTTGGACACAGTGCATTTGTTCTTGCTGATCCAGCAATCAGAGCTCCTG 240
QY 1725 AGAAACTCTACCAAGAGAGAGAGAGTGCCTCTCTGAGGAGAGAGAGAGAGAGAGTCACTC 1784
Db 241 AAAGTCTTTAACTGGACCAAGAGAGAGAAATACCTGAAGAGAAACATCCAAAATGTTCGA 300
QY 1785 GGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGAAAAATTTGAACCTGCACT 1844
Db 301 AACGCATTCGGAAGTATGCCGAGGAGGTGAAAGTGAATGGGATAAGCTGAGCAATCGCT 360
QY 1845 CCGCTGACTGGCAGAGAGAAAAATAGATGAGACCCCTTGAAGAGACTCCAGGAACTTCAAGAGG 1904
Db 361 CTGTTGATGGCAGAGAGCGTATAGATGAGGCCTTGAAGAGACTTGAAGATTTGCAAGATT 420
QY 1905 CCACGGATGAGCTGGACCTCAAGCTGGCCCAAGCTGAGGTGATCAAGGGATCCTGGCAGC 1964
Db 421 CAATGGATGAATTAACCTCAAAATTTGAGACAGGCTGAAGCTATCAAGATACATGGCAAC 480
QY 1965 CCGTGGCGGATCTCCTCATTTGACTCTCTCAAGATCACCTCGAGAGAAAGTCAAGGCACTTC 2024
Db 481 CTGTCGGGGATCTACTAATAGACTCATTTGACGATCATATTGAAAAAGTCAAGGTTTTC 540
QY 2025 GAGGAGAAATTCGCGCTCTGAAAGAGAAACGCTGAGCCACGTCATGACCTTGTCTGCCAGC 2084

Db 541 GAGCAGAAATTCCTCCATGAAGGAAATGTGACTCACATGAATGATCTCGTCTCCAGT 600
QY 2085 TTACCACTTTTGGGCAATTCAGCTCTCACCGTATACCTCAGCACTCTGGAAGACTGAACA 2144
Db 601 TCACACCACCTGATATCAATATCCCGTACAAATTAACCAACAGTTGGAGGATCTGAACA 660
QY 2145 CCAGATGGAAGCTTCTGAGGTGGCGCTCGAGGAGCCGAGTCAGGCAGCTGCATGAAGCC 2204
Db 661 CACGGTGGAAACTCTTGACGATCTATAGATGAGCTTCTGAAAGCAGCTGCATGAGGCTC 720
QY 2205 ACAGGACTTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGCTGTCTCCAGGCTCCCTGG 2264
Db 721 ATAGAGATTTTGGACCAACATCCAGCACTTCTTATCAACTTCTGCAAGGCTCTTGGG 780
QY 2265 AGAGAGCATCTCGCCAAACAAAGTGCCTACTATATCAACCCACGAGACTCAAAACAATT 2324
Db 781 AGCGAGCAATATCAACCAACAAAGGTTCCCTACTACATTAACCAACCAACTCAGACCCT 840
QY 2325 GCTGGGACCATCCCAAAATGACAGAGCTTCTTACCAGCTTTTAGCTGACCTGAATAATGTC 2384
Db 841 GCTGGGATCACCCCAAGATGACGAGCTCTATCAGTCACTAGCCGATCTAAATAATGTC 900
QY 2385 GATTCTCAGCTTATAGGACTGCCATGAAACCTCCGAGACTGCAGAAAGCTCTTGTCTGG 2444
Db 901 GGTCTTCTGCATATAGAACTGCCATGAAACTGCGAGACTGCAGAAAGCTCTTGTCTGG 960
QY 2445 ATCTCTTGTAGCCTGTGAGCTGCATGATGCTTGGACCAACCACTCAAGCAAAATG 2504
Db 961 ATCTCTCTCCCTGCCAAGTGTGTGAGGCTTTGTGAGGCTTTGTGACGACCACTGAAACAG 1020
QY 2505 ACCAGCCCATGGATATCCTGCAGATTAATTTGTTGACCACTATTTATGACCGCTGG 2564
Db 1021 ATCAACTGTGGATATTTTAGAAATAATACTGCTTGAATCAATCTATGATCGGCTGG 1080
QY 2565 AGCAAGAGCACAACAATTTGGTCAACGCTCCCTCTCTGCGTGGATATGTCTGAAGTGGC 2624
Db 1081 AGCAGGAACACAGCAACCTTTGTCAACGCTGCCTCTTTGTGGACATGTGTTAAACTGGC 1140
QY 2625 TGCTGAATGTTATGATACGGGACGACAGGAGGAGGATCCGTGTCTCTCTTTTAAACTG 2684
Db 1141 TGCTCAACGCTATGACACTGGTGCAGAACGAGGAAGATCCGTGTCTCTCTTAAAGTG 1200
QY 2685 GCATCATTTCCCTGTGTAAAGCAGATTTGGAAGACAGTACAGATACCTTTTCAAGCAAG 2744
Db 1201 GAATCATGTCTATGTGCAAAAGCACACCTGGAGGACAAATACCGTTATCTGTTCAAGCAAG 1260
QY 2745 TGGCAAGTTCAACAGGA 2761
Db 1261 TGGCGAGTCCCACTGGA 1277

RESULT 7

MUSDYS MUSDYS 3275 bp mRNA linear ROD 27-APR-1993
LOCUS Mouse dystrophin mRNA, partial cds.
DEFINITION M18025
ACCESSION M18025
VERSION M18025.1 GI:192971
KEYWORDS dystrophin.
SOURCE Mouse adult heart, cDNA to mRNA.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
REFERENCE 1 (bases 1 to 3275)
AUTHORS Hoffman, E.P., Monaco, A.P., Feener, C.C. and Kunkel, L.M.
TITLE Conservation of the Duchenne muscular dystrophy gene in mice and humans
JOURNAL Science 238 (4825), 347-350 (1987)
MEDLINE 88018015
PUBMED 3659917
FEATURES Location/Qualifiers
1. .3275
/organism="Mus musculus"

BASE COUNT 887 a 784 c 802 g 688 t
ORIGIN
RWKLLQVSVDDRLKQLQEAHRDFGPPSSQHFSTSVLPWQRSISHNKVPYYINHQTQT
TCWDHPKMTLEFQSLADLNNVRESAYRTAIKIRRLQKALCLDLDLELNTTNEVFKQHKL
NNDOLLSPDVINCLTYYDGLQHLKDLVNPVPLQVDMCLNWLNNVYDTGRTGKIRV
QSLKIGLSLGLLEEKYRCLKEVAGPTQEMCDQRLGLLHDAIQIPROLGEVAAF
GGSNIEPSVRSFQNNKPKETSVKEFIDWMLHLEPQSMVWLPVLRVAAETAKHQAK
CNICKECPVIGFRYSLKHFNDVQCSCFFSGRTAKGHLKHPMVEYCIPTTSGEDVR
DFTKVLNKFERSKKEVAKHPRGLYLPVQTVLGGDNLETPIITLSMWPEHYDPSQSPQL
PHDDTHSRIEQATRLAOMERINGSLTDSSTGSEVEDEHALIQYQCTLGGEPSVS
QPQSPAQILKSVEREERLELRIADLEEEQRLQVEYEQLEKHLRGLPLVGSPPDS
IVSPHHTSELSIAEAKLLRQKRLRLEARMQILEDHKNQLESQHLRLQLLEQPDSD
SRINGVSPWASPOHSALSYSLDTPDQPFQHAASEDLLAPPHDTSTDLTDVMEQINST
FPSCSNVPSRPQAM"

Query Match	Score	DB	Length	Indels	Mismatches	Gaps
Best Local Similarity	21.9%	604.8	10	3161	532	9
Matches	943	Conservative	0			
1285	GAAACAATGGCAAGACCTCCAAAGTGAAATGAAGCTCACACAGATGTTTATCACAACT	1344				
329	GAAGCTCTGCAAGGACATCCAGGCAGAAATGATGCCCAATGACATATTTAAAGCAT	388				
1345	GGATGAAACAGCCAAATAATCCTGAGATCCCTGGAAAGTTCCGATGATGACAGTCTCTGTT	1404				
389	CGATGGAACCCGGCAGAAAGATGGTGAAGCTCTGGGAAATCTGAGGAGCAACAATGCT	448				
1405	ACAAAGACGTTGGATAACATGAATCTCAAGTGGAGTGAATTCGGAAAAAGTCTCTCAA	1464				
449	TCAGCATCGACTGGATGACATGAACCAAGATGGAATGATTGAAGGCAAAATCTGCTAG	508				
1465	CATTAGGTCCTTGGAAAGCCAGTTCTGACCAAGTGAAGCGTCTGCACCTTTCTCTGCA	1524				
509	CATCAGGGCCCATTTGGAGGCCAGTGTGAGAAATGAGAAATGAGAAATGAGAAATGAG	568				
1525	GGAACCTTCTGGTGTGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTG	1584				
569	AGAGCTGATCAAAATGGCTCAATATGAAAGATGAGGAGCTTAAGAAAGCAGATGCCCAT	628				
1585	AGCGGACTTTCAGCAGCTTCAAGAACGACGATGTACATAGGCGCTTCAAGAGGGAATT	1644				
629	AGGGGACGTCCTGCTTACAGCTTCCAGTATGACCACTGCAAGGTGCTGAGACGTGAGCT	688				
1645	GAAACTAAAGAACCTGTAAATCATGATGACTCTTGAGACTGTACGAATATTTCTGACAGA	1704				
689	AAAGGAGAAAGAGTATTCTGTGCTGAACCCGCTAGATCAAGCTCGAGTTTCTGGCTGA	748				
1705	GCAGCCTT-----GGAAGGACTAGAGAACTCTACCAGGAGCCCGCAGAGAGCTGCC	1755				
749	TCAGCCCAATAGAGGCCCCCGAAGAACCAAGAAACCCACAAATCAAGAACAGAGATTGAC	808				
1756	TCCTGAGGAGAGAGCCCGCAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGAGTCAA	1815				
809	TCCTGAGGAGAGAGCCCGCAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTTCTGAGTCCG	868				
1816	TACTGAGTGGGAAAAATTAACCTGCACCTCCGCTGCTGACGAGAAAAATAGATGAGAC	1875				
869	AGAGAAGTGGGAAAAATTAACCTGCACCTCCGCTGCTGACGAGAAAAATAGAGAGGC	928				
1876	CCTTGAAGACTCCAGGAACTTCAAGAGGCCACCGATGAGCTGGACCTCAAGCTGCGCCA	1935				
929	GTTAGAGAAACTCCGAGACCTGCGAGGAGCTATGGACGACCTGGACGACATGAAGGA	988				
1936	ACCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGGCGATCTCCTCATTTGACTCTCCA	1995				
989	GGTGAGGCTGTGCGGAATGGCTGGAAGCCCGTGGGAGACCTGCTTATAGACTCCCTGCA	1048				
1996	AGATCACCTCGAGAAAGTCAAGGCACTTCGAGGAGAAATTTGGCCCTCTGAAGAGAACGT	2055				
1049	GGATCACATCGAGAAACCCCTGGCGTTTAGAGAGAAATTTGACCAATCACTTAAAGT	1108				
2056	GAGCCACGTCATGACCTTGTCTCGCCAGCTTACCACCTTTGGGCAATTCAGCTCTCACCGTA	2115				

Db	1109	AAAAACAATGAATGACCTGTCCAGTCAGCTGTCTCCACTTGACTTGACATCCATCTCTAAA	1168
QY	2116	TAACCTCAGCACTCTGGAAGACCTGAACACACAGATGGAAGCTTCTGCAGGTGGCCGTCGA	2175
Db	1169	GATGCTCCGACGCTGGATGACCTTAATATGCGATGGAAGCTTCTACAGGTTTCCGTCGA	1228
QY	2176	GGACCGAGTCAGGCAGCTGCATGAAGCCACAGGAGCTTTGGTCCAGCATCTCAGCACTT	2235
Db	1229	CGATCGCCTTAAGCAGCTCCAGGAAGCCACAGAGATTTGGGCCATCTTCTCAACACTT	1288
QY	2236	TCCTTCCACGCTCTCCAGGCTCCCTGGGAGAGAGGCACTCTCGCCAAACAAAGTGCCCTA	2295
Db	1289	TCTGTCCACTTCAGTCCAGCTGCCGTGGCAGAGATCCATTTCACATAATAAAGTGCCCTA	1348
QY	2296	CTATATCAACACGAGACTCAAAACAACTTGCTGGGACCAATCCCAAAATGACAGAGCTCTA	2355
Db	1349	TTACATCAACCATCAACACAGACAAACCTTTGGGATCATCTCTAAATGACTGAGCTCTT	1408
QY	2356	CCAGTCTTTAGCTGACCTGAATAATGTCAAGATTTCTCAGCTTATAGGACTGCCATGAACT	2415
Db	1409	CCAATCCCTTGTGATCTGAATAATGTACGTTTCTCTGCTACCGCAGCAATCAAAAT	1468
QY	2416	CCGAAGACTGCAGAAGGCCCTTTGCTTGGATCTCTTGGAGCTGTGAGCTGCATGTGATGC	2475
Db	1469	TCGAAGCTGCAAAAGCATTTATGCTGGATCTCTTAGAGCTGAATACGACGAATGAAGT	1528
QY	2476	CTTGGACCAAGCAACCTCAAGCAAAATGACCAAGAGAGCAACAATTTGGTCAACGTCCTC	2535
Db	1529	TTTCAAGCAGCACAACCTGAACCAAAATGATGATGATGATGATGATGATGATGATGATG	1588
QY	2536	TTGTTTGGACCACTATTATGACCGCTGGAGCAAGAGAGCAACAATTTGGTCAACGTCCTC	2595
Db	1589	CTGCTGACCACTTACGATGGCTTGAGCAGCTGCACAGCTGCACAGCTGCATCAATGTTCC	1648
QY	2596	TCTCTGCGTGGATATGTCTGAACTGGCTGGTGAATGTTATGATACGGGACGAACAGG	2655
Db	1649	ACTCTGCGTGCATATGTCTCAACTGGCTGCTCAACGATACGACACGGCCGAGCTGG	1708
QY	2656	GAGGATCCGTGCTCTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGA	2715
Db	1709	AAAAATTCGGGTACAGAGCTGCTGAAGATTTGATGATGCTCTCTCAAAAGGCTCTTTA	1768
QY	2716	AGACAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAG	2759
Db	1769	AGACAATACAGATGCTCTTTTAAAGAGGTTGGCAGGCGGCAACTG	1812

RESULT 9
E30223
LOCUS Shortened dystrophin.
DEFINITION E30223
ACCESSION E30223
VERSION JP 1999318467-A/6.
KEYWORDS unclassified.
SOURCE unclassified
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 3163)
AUTHORS Sinichi,T.
TITLE Shortened dystrophin
JOURNAL Patent: JP 1999318467-A 6 24-NOV-1999;
SCIENCE & TECH AGENCY,NATIONAL CENTER OF NEUROLOGY AND PSYCHIATRY
COMMENT OS Unidentified
PN JP 1999318467-A/6
PD 24-NOV-1999
PF 08-MAY-1998 JP 1998142134
PR
PI SINICHI TAKEDA
PC C12N15/09,A61K48/00,C12N15/00
CC Strandedness: Both;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1. .3163


```
FEATURES             Source
FT                  Location/Qualifiers
                   1..3163
                   /organism="Unidentified".
BASE COUNT      930 a 731 c 720 g 782 t
ORIGIN
Query Match      20.5%; Score 565.6; DB 6; Length 3163;
Best Local Similarity 98.4%; Pred. No. 4.2e-135;
Matches 571; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY 2182 AGTCAGGCAGCTGCATGAAGCCACAGGGACTTTGGTCCAGCATCTCAGCACTTTCITTC 2241
Db 942 AGCCATCCAGGAAGTGGAAGCCACAGGGACTTTGGTCCAGCATCTCAGCACTTTCITTC 1001
QY 2242 CACGTCTGTCCAGGTCCTGGGAGAGAGCCATCTCGCCAAACAAAGTCCCTACTATAT 2301
Db 1002 CACGTCTGTCCAGGTCCTGGGAGAGAGCCATCTCGCCAAACAAAGTCCCTACTATAT 1061
QY 2302 CAACCACGAGACTCAACAACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAAGTC 2361
Db 1062 CAACCACGAGACTCAACAACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAAGTC 1121
QY 2362 TTTAGCTGACCTGAATAATGTCAGATCTCAGACTTATAGGACTGCCATGAACTCCGAAG 2421
Db 1122 TTTAGCTGACCTGAATAATGTCAGATCTCAGACTTATAGGACTGCCATGAACTCCGAAG 1181
QY 2422 ACTGCAGAAAGCCCTTTGCTTGGATCTCTTGAGCCTGTGCAGCTGCATGTGATGCCCTCTG 2601
Db 1182 ACTGCAGAAAGCCCTTTGCTTGGATCTCTTGAGCCTGTGCAGCTGCATGTGATGCCCTCTG 1301
QY 2482 CCAGCACAACTCAAGCAAAATGACCAAGAGAGCAACAACTTTGGTCAACGTCCCTCTCTG 2541
Db 1242 CCAGCACAACTCAAGCAAAATGACCAAGAGAGCAACAACTTTGGTCAACGTCCCTCTCTG 1361
QY 2542 GACCACTATTTATGACCGCCTGGAGCAAGAGCAACAACTTTGGTCAACGTCCCTCTCTG 2661
Db 1302 GACCACTATTTATGACCGCCTGGAGCAAGAGCAACAACTTTGGTCAACGTCCCTCTCTG 1421
QY 2602 CGTGGATATGTGTCTGAAGTGGTCAACGTCCCTCTCTG 2761
Db 1362 CGTGGATATGTGTCTGAAGTGGTCAACGTCCCTCTCTG 1481
QY 2662 CGTGTCTCTCTTTTAAACTGGCATATTCCTGTGTAAAGCACATTTGGAAGACAA 1521
Db 1422 CGTGTCTCTCTTTTAAACTGGCATATTCCTGTGTAAAGCACATTTGGAAGACAA 1521
QY 2722 GTACAGATACCTTTTCAAGCAAGTGGCAAGTTCACACAGGA 2761
Db 1482 GTACAGATACCTTTTCAAGCAAGTGGCAAGTTCACACAGGA 1521
RESULT 10
AF195788
LOCUS
DEFINITION
Rattus norvegicus 3073 bp mRNA linear ROD 02-MAY-2001
variant (Drp2) mRNA, complete cds.
ACCESSION
AF195788
VERSION
KEYWORDS
SOURCE
Rattus norvegicus.
ORGANISM
Rattus norvegicus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
EFERENCE
1 (bases 1 to 3073)
AUTHORS
Roberts,R.G. and Sheng,M.
TITLE
Association of dystrophin-related protein 2 (DRP2) with
postsynaptic densities in rat brain
JOURNAL
Mol. Cell. Neurosci. 16 (5), 674-685 (2000)
MEDLINE
20538711
PUBMED
11083927
```

```
REFERENCE           2 (bases 1 to 3073)
AUTHORS             Roberts,R.G. and Sheng,M.
TITLE               Direct Submission
JOURNAL             Submitted (18-OCT-1999) Division of Medical and Molecular Genetics,
                    GKT Medical School, 8th Floor, Guy's Hospital, Guy's Tower, London
                    SE1 9RT, UK
FEATURES             Location/Qualifiers
                    source
                    1..3073
                    /organism="Rattus norvegicus"
                    /strain="Sprague-Dawley"
                    /db_xref="taxon:10116"
                    1..3073
                    /gene="Drp2"
                    69..70
                    /gene="Drp2"
                    /note="Drp2"
                    /note="splice acceptor AG dinucleotide used by A-form"
                    200..3073
                    /gene="Drp2"
                    /note="DRP2; membrane-associated cytoskeletal protein;
                    contains two spectrin repeats; WW domain; zz domain"
                    /codon_start=1
                    /product="dystrophin-related protein 2 B-form splice
                    variant"
                    /protein_id="AAG28485.1"
                    /db_xref="GI:11066167"
                    /translation="MQPLVMQGCPTLPRCHEWHAADRHHSSSLRNTCPQPVRAAV
                    TIPAPPWDGADGPCLSPKLLNGSVGAVGLEPSAMNLWNKKEKSHNLRLEAFSD
                    HSGKLQPLQELIDWLSQDEELSAQLPLQGDVALVQKEKETHAAFMEEVKSGPYIY
                    SVLESAQAFLSQHPFEELESHSESKDTSRQRIQNLRFVWKQATVASELWEKLTAR
                    CVDQHRHIEHTLHLELQIGAMEELSTLQAEGRVATWEPIDGLFIDSLPEHIQAIK
                    LFKEEFPVKDGVKLVNDLAHQLAISDVHLTSMENSRALAEQINVRWKQLQVSVAEERLQ
                    LQDAHRDFGSGHFLSTSVQVPWERAIQVPRQLGEVAAFGGSNVESVRSRFR
                    ADLNNIKFSAYRTAMKLRVQKALRLDLVLTLLTLEIFNEHDLQASEHMDVVEVIHC
                    LTALYERLEERGILNVPLCVDMSLNLLNVFDSGRSGKMRALSFKTGIAICLCGTEV
                    KEKLQVLFQVANSKSCDQRHLGALLHEAIQVPRQLGEVAAFGGSNVESVRSRFR
                    STGKPVIEASQFLEWVNLQPSQMVWLAVLHRTVAEQVKHQTCSICRQCPIKGFYR
                    SLKQFNVDICOTFLTGRASGKNGKHYPIMEYVPTTSSENMRDEATTLKKNFRSKQY
                    FSKHPQRGYLPVQSVLESDCSETPASSPMLPHADTHSRIEHFASRLAEMESQNSFFN
                    DLSLDDSIDEDQYLLRHSSPITDRPAFQQAQPCSMATESKGELEKILAHLEDENRI
                    LQELRLKWOHEEAVEAPTLAGESAEATPDHNEELAEARILRQHKSRLETRMQIL
                    EDHNKQLESQQLRLELLQPPTESDNGSAGSSSLASSPROSGSHPREKGQTPDTE
                    AADDVGSKSQDVSHSLEDIMEKLRHAFPSVRSDDVTANTLLAS"
BASE COUNT      779 a 843 c 790 g 659 t 2 others
ORIGIN
```

```
Query Match      19.0%; Score 523.4; DB 10; Length 3073;
Best Local Similarity 62.0%; Pred. No. 3.5e-124;
Matches 827; Conservative 0; Mismatches 506; Indels 0; Gaps 0;
QY 1423 CATGAACCTTCAAGTGGAGTGAACCTTCGAAAAAGTCTCTCAACATTAGTCCCATTTGGA 1482
Db 433 CATGAATCTGTGTGGAATGAAATAAAAAAGAGTCTCACAACTCCGCGCTCGCCTAGA 492
QY 1483 AGCCAGTCTGACCCAGTGAAGCGTCTGCACCTTTCTCTGCAGGAAGTCTTGGTGGCT 1542
Db 493 GGCCTTCTCAGACCCACAGTGGAAAACTTCAGCTCCACITTCAGAGATTATTGACTGGCT 552
QY 1543 ACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTGGAGGGCAGTTCAGCAGT 1602
Db 553 CAGCCAAAAGGATGAAGAATTTGACGCTCAGCTGCCCTTCAAGGGGATGTGGCCCTGGT 612
QY 1603 TCAGAAAGCAGACGATGTACATAGGGCCCTTCAGAGGGGAATTTGAAACCTAAAGAACCTGT 1662
Db 613 ACAACAGGAGAGAGACACATGCTGCCCTTTATGGAAGAGTGAAGTCTAAAGGCCCTA 672
QY 1663 AATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAAGACT 1722
Db 673 TAICTACTCTGTCTGGAATCGGCACAGAGCTTTCCTGTCTCAGCACCCATTTGAGGAATT 732
QY 1723 AGAGAAACTCTACAGGAGCCCGCAGAGAGCTGCCTCTCTGAGGAGAGAGCCAGCAATGTAC 1782
Db 733 AGAGGAGTCTCATTTCTGAGAGCAAGATACCTCCCCCAAGACAGCGGATTCAAAACCTCAG 792
```


variant (Drp2) mRNA, complete cds.	
AF195787	AF195787.1 GI:11066164
VERSION	
KEYWORDS	
SOURCE	Rattus norvegicus.
ORGANISM	Rattus norvegicus
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
	Rattus.
REFERENCE	1 (bases 1 to 3140)
AUTHORS	Roberts, R.G. and Sheng, M.
TITLE	Association of dystrophin-related protein 2 (DRP2) with postsynaptic densities in rat brain
JOURNAL	Mol. Cell. Neurosci. 16 (5), 674-685 (2000)
MEDLINE	20538711
PUBMED	11083927
REFERENCE	2 (bases 1 to 3140)
AUTHORS	Roberts, R.G. and Sheng, M.
TITLE	Direct Submision
JOURNAL	Submitted (18-OCT-1999) Division of Medical and Molecular Genetics, GKT Medical School, 8th Floor, Guy's Hospital, Guy's Tower, London SE1 9RT, UK
FEATURES	Location/Qualifiers
source	1..3140
	/organism="Rattus norvegicus"
	/strain="Sprague-Dawley"
	/db_xref="taxon:10116"
gene	1..3140
	/gene="Drp2"
CDS	267..3140
	/gene="Drp2"
	/note="DRP2; membrane-associated cytoskeletal protein; contains two spectrin repeats; WW domain; ZZ domain"
	/codon_start=1
	/product="dystrophin-related protein 2 A-form splice variant"
	/protein_id="AAG28484.1"
	/db_xref="GI:11066165"
	/translation="MQPLVMQGCPTLPRCHEHWAADRFHSSSLRNTCPQPVRAAV TIPAPPWDGAGDPCLSPKLLNGSVGAVGLEPSAMNLCWNEIKKSHNLRLEAFSD HSKLQPLQEIIDLQSKDELSAQLPLQGDVALVQOEKETHAAFMEEVKSGPYI SVLESAQAFLSQHPFEELESHESKDTSPQRIONLSRFVWKQATVASSELWEKLTAR CVDQHRHIEHTLEHLEIQAMEELSSLTQAEGRATWEPIDGLFIDSLPEHIOAIK LFKFEESPVKDGKLVNDLAHQLAISDVHLSMENSRALEQINVRWKQLQVSVARLKQ LODAHRDFGPGSQHFLSTSVQVPERAISPNKVPYIINHQATTCWDHPKMTELYOTL LTALYERLEEEERILNVPLCVDMSLNLLNLFDSGRSGKMRALSFKTGIACLCGTEV KEKLYLFQSVANSKQDQRLHGLALLHEAIOVPROLGEVAAPFGSNVPSVRSQFRY STGKPVIEASQFLEWVNLQPSVMVLAHLRVYTPPTSSSNMRDFTATLKNKRSKQY SLKQENVDICQTCFLTGRASKGNKHLHYPIMEYTPPTSSSNMRDFTATLKNKRSKQY FSKHPDQSIDEDQYLLRHSSPITDREPAPGQQAQPCSMATESKGELEKILAHLEDENRI DLSLSPDQSIDEDQYLLRHSSPITDREPAPGQQAQPCSMATESKGELEKILAHLEDENRI LQELRRLLKQHEEAVEAPTLAGESAETPDHNEELAEARILROHKSRLTRMOIL EDHNKQLESQQLRLRELLQPPTESDNGSGAGSSLASPPROSEGSHPREKQGTTPDTE AADDVGSKSQDVSHSLEDIMEKLRHAFPSVRSDDVTANTLLAS"
BASE COUNT	790 a 864 c 794 g 690 t
ORIGIN	
Query Match	19.0%; Score 523.4; DB 10; Length 3140;
Best Local Similarity	62.0%; Pred. No. 3.5e-124;
Matches	827; Conservative 0; Mismatches 506; Indels 0; Gaps 0;
QY	1423 CATGAACCTCAAGTGGAGTGAACCTCGGAAAAAGTCTCTCAACATTAGTCCCATTTGGA 1482
Db	
QY	1483 AGCCAGTTCTGACCGAGGAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGTGGCT 1542
Db	
QY	1543 ACAGCTGAAAGATGATGAATTAAGCCGGCAGGACCTATTTGAGGGCGACTTTCCAGCAGT 1602
Db	
QY	1603 CAGCCAAAGGATGAAGAATTGTTCAGCTCAGCTGCCCCTTGCAGGGGATGTGGCCCTGGT 679
Db	
1783	TCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGAAAAATTTGAACCTGCA 1842
1793	TCGGCTTGTGTGAAGCAGGCAACAGTGGCCAGTGAACCTGTGGGAGAAAGCTGACAGCCG 852
1843	CTCCGCTGACTGGCAGAGAAAAATATAGATGAGACCCCTTGAAGACTCCAGGAACCTTCAAGA 1902
853	CTGTGTGGATCAGCATCGCCACATTGAGACACACCCTGGAACATCTATTGGAGATCCAAGG 912
1903	GGCCACGGATGAGCTGGACCTCAAGCTGGCCCAAGCTGAGGTGATCAAGGGATCCTGGCA 1962
913	GGCAATGGAGAACTGAGCAGTACTTTGACCAAGCAGAGGGAGTCCGAGCCACATGGGA 972
1963	GCCCGTGGGGATCTCCTCATTTGACTCTCTCCAGATCACCTCGAGAAAGTCAAGGCACT 2022
973	GCCATAGGAGATCTCTTATCGATTCCCTCCAGAGCATATCCAAGCCATCAAGTTATT 1032
2023	TCGAGGAGAAATTCGCCCTCTCTGAAAGAACCTGAGCCACGTCATGACCTTGTCTGCCCA 2082
1033	CAAGAAGAATTTCTCTCCTGTGAAGACGGGTGAAGTTAGTGAATGACCTGGCCCAACA 1092
2083	GCTTACCACCTTTGGGCATTCAGCTCTCACCGTATACCTCAGCACTCTGGAAGACCTGAA 2142
1093	GCTTGTCTATTCTGTATGTCACCTTCAATGGAGAAATTCAGGGCTCTGGAACAGATCAA 1152
2143	CACCAGATGGAAGCTTCTGCAGGTGGCGCTCGAGGACCGAGTCAAGCAGCTGCATGAAGC 2202
1153	CGTCCGCTGGAACAGCTCCAGGTGTCAGTTGCTGAGAGGCTTAACACACCGCTCAGACCAC 1212
2203	CCACAGGGACTTTGGTCCAGCATCTCAGCACTTCTTCCACGCTGTGTCAGGGTCCCTG 2262
1213	CCACCGGGACTTTGGGCTGGGTCCAGCACTTCTCTCCACTTCTGTTCAGTTCCCTG 1272
2263	GGAGAGAGCCATCTCGCAAAACAAAGTGCCTACTATATCAACACCGAGACTCAACAAAC 2322
1273	GGAAAGAGCAATTTCTCCCAATAAAGTCCCTACTACATCAACACCGAGGCTCAGACCAC 1332
2323	TTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAAGTCTTTAGCTGACCTGAATATGT 2382
1333	ATGCTGGGACCATCCCAAGATGACCGAGTTATACCAAAACCTAGCTGACCTGAACACAT 1392
2383	CAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAAGACTGCAAGAGGCCCTTTGCTT 2442
1393	CAAGTTCTCAGCTTACCGCACTGCCATGAAGCTCCGCCGAGTCCAGAAAGGCCCTGCGCTT 1452
2443	GGATCTCTTGAAGCTGTCAGCTGCATGTGATGCCTTGGACCAAGCAACACCTCAAGCAAAA 2502
1453	GGATCTGGTAACCTTAACATACAGCTCTGGAGATCTTCAATGAGCAGCAGCTTGCAGGCCAG 1512
2503	TGACCAAGCCCATGGATATCCTGCAGATTATTATTTGTTGACCACTATTTATGACCGCCT 2562
1513	TGAACATGTAATGGATGTGGTGGAGGTCTATTCAGCTGCTGACTGCCTTGTATGACGACT 1572
2563	GGAGCAAGAGCACAACAATTTGGTCAACGCTCCCTCTCTGGTGGATATGTGTCTGAACTG 2622
1573	GGAGGAGGAAAGAGGCATCTCTGGTCAACGCTGCCGCTGTGTGTAGACATGAGCCTCAACTG 1632
2623	GCTGCTGAATGTTTATGATACGGGACGAGGAGGATCCGCTGCTCTTTTAAAC 2682
1633	GCTCCTCAACGTTTGTATAGTGGCCGAGTGGAAAGATGCGAGCATTTGCTCTTAAGAC 1692
2683	TGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAAGTACAGATACCTTTTCAAGCA 2742
1693	TGGCATCGCATGCTGTGTGGCACTGAAGTGAAGAAAGAAACTTTCAGTATCTCTTCAGCCA 1752
2743	AGTGGCAAGTTCA 2755
1753	AGTGGCAATTTCA 1765
RESULT 11	
AF195787	3140 bp mRNA linear ROD 02-MAY-2001
LOCUS	
DEFINITION	Rattus norvegicus dystrophin-related protein 2 A-form splice

QY 1603 TCAGAAGCAGAACGATGTACATACGGCCCTTCAAGAGGGGAATTGAAAACTAAAGAACCTGT 1662
Db 680 ACAACAGGAGAGGAGACATGCTGCCTTTATGGAAGAGTGAAGTCTAAAGGCCCTTA 739
QY 1663 AATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGAAGGACT 1722
Db 740 TATCTACTCTGTGGAATCGGCACAGGCTTTCTCTCTCAGCACCTTGTAGGAAT 799
QY 1723 AGAGAACTCTACAGAGCCAGAGAGCTGCCCTCTCTGAGGAGAGAGCCAGAAATGTAC 1782
Db 800 AGAGGAGTCTCACTCTGAGAGCAAGATACCTCCCAAGACAGCGGATTCAAAACCTCAG 859
QY 1783 TCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGAAAAATTTGAACCTGCA 1842
Db 860 TCGCTTTGTGGAAGCAGGCAACAGTGGCCAGTGAAGTGTGGGAGAGCTGACAGCCCG 919
QY 1843 CTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAAGACTCCAGGAACTTCAAGA 1902
Db 920 CTGTGTGGATCAGCATCGCCACATTGACACACCCCTGGAACATCTATTGGAGATCCAAGG 979
QY 1903 GGCCACGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGATCCTGGCA 1962
Db 980 GGCAATGGAGAACTGAGCAGTACTTTGACCCCAAGCAGAGGGAGTCCGACCATGGGA 1039
QY 1963 GCGCGTGGCGGATCTCCTCAATGACTCTCTCAAGATCACTCGAGAAAGTCAAGGCAC 2022
Db 1040 GCCCATAGGAGATCTCTTTATCGATTCCCTCCAGAGCATATCCAAAGCCATCAAGTTAT 1099
QY 2023 TCGAGGAGAAATTTGGCCCTCTGAAAGAGAACGTCGAGCCACGTCATGACCTTGCTCGCA 2082
Db 1100 CAAAGAAGAAATTTCTCTCTCTGAAAGACGGGGTGAAGTTAGTGAATGACCTGGCCCCA 1159
QY 2083 GCTTACCACCTTTGGGCAATCAGCTCTCAAGTATAACCTCAGCAGCTCTGGAAGACCTGAA 2142
Db 1160 GCTTGCTATTTCTGATGTGCACCTTGTCAATGGAGAATTCAGGGCTCTGGAACAGATCAA 1219
QY 2143 CACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACGAGTCAGGACCTGACCTTGCTCGCA 2202
Db 1220 CGTCCGCTGGAACAGCTCCAGGTGTGAGTGTGAGAGGCTTAACACAGCTCCAGGATGC 1279
QY 2203 CCACAGGGACTTTGGTCCAGCATCTCAGCAGCTTCTTCCAGCTCTGTCAGGGTCCCTG 2262
Db 1280 CCACCGGACTTTGGGCTTGGTTCACAGCAGCTTCTCTCCACTTCTGTTCAAGTTCCTGT 1339
QY 2263 GGAGAGAGCCATCTCGCCCAACAAAGTGCCTACTATATCAACCCAGCAGACTCAAAAC 2322
Db 1340 GGAAGAGCAATTTCTCCCAATAAAGTCCCTACTATACATCAACCCAGGCTCAGACCCAC 1399
QY 2323 TTGCTGGGACCATCCCAAAATGACAGAGCTCTACAGTCTTTAGCTGACCTGAATAATGT 2382
Db 1400 ATGCTGGGACCATCCCAAGATGACCGAGTATACCAAAACCTAGCTGACCTGAACAAAT 1459
QY 2383 CAGATTCTCAGCTTATAGGACTGCAATGAAACTCCGAAGACTGCAGAGGCGCCCTTTGCTT 2442
Db 1460 CAAGTTCTCAGCTTACCGCACTGCAATGAAAGTCCCGGAGTCCAGAGGCGCCCTGCGCTT 1519
QY 2443 GGATCTCTTGAGCCTGTGAGTGCATGTGATGCCTTGGACCCAGCACAACCTCAAGCAAA 2502
Db 1520 GGATCTGGTAACCTTTAATACATGACAGCTCTGGAGATCTTCAATGAGCAGCAGCTTGCAGGCCAG 1579
QY 2503 TGACCAAGCCATGGATATCTGAGATTTGTTCAAGCTCCCTCTGCGTGGATATGTGCTGAAC 2562
Db 1580 TGAACATGTAATGGATGGTGGAGGTCAATTCAGTCTGCTGACTGCTTGTATGAGCAGCT 1639
QY 2563 GGAGCAAGAGCACAACAATTTGGTCAAGCTCCCTCTGCGTGGATATGTGCTGAAC 2622
Db 1640 GGAGGAGGAAAGAGCAGCTCTGTTCAACCTGCGGCTGTGTAGACATGAGCCTCAACTG 1699
QY 2623 GCTGCTGAATGTTATGATACGGGACGAAACAGGGAGGATCCGTTGCTGCTTTTAAAC 2682
Db 1700 GCTCCTCAACGTTTGTAGTGGCCCGCAGTGGAAAGATCGGAGCATTTGCTCTTAAGAC 1759

QY 2683 TGGCATCATTTCCCTGTGTAAAGCACACATTTGGAAGACAAGTACAGATACCTTTTCAAGCA 2742
Db 1760 TGGCATCGCATGCCTGTGTGGCACTGAAGTGAAGAAAACTTCAGTATCTTTCAGCCA 1819
QY 2743 AGTGGCAAGTTCA 2755
Db 1820 AGTGGCCAATTCA 1832
RESULT 12
HSU43519
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
AUTHORS
1 (bases 263 to 3189)
Roberts,R.G., Freeman,T.C., Kendall,E., Vetrie,D.L., Dixon,A.K.,
Shaw-Smith,C., Bone,Q. and Bobrow,M.
TITLE
JOURNAL
MEDLINE
PUBMED
96225452
REFERENCE
2 (bases 1 to 3499)
Roberts,R.G.
Direct Submission
Submitted (16-DEC-1995) Roland G. Roberts, Medical Genetics,
Addenbrooke's Hospital, 3rd Floor, Lab Block, Addenbrooke's
Hospital, Hills Road, Cambridge, Cambs CB2 2Qq, UK
FEATURES
source
1. .3499
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="Xq22"
1. .3499
/gene="DRP2"
/gene="DRP2"
/gene="DRP2"
/gene="DRP2"
/gene="DRP2"
/function="membrane cytoskeleton"
/codon_start=1
/product="dystrophin-related protein 2"
/protein_id="AAC50538.1"
/db_xref="GI:1353782"
/translation="MVMQGCPTLPRCHDWQAADQFHSSSLRSTCPHPQVRAAVTSP
APQDAGVPCLSKLKLVSVGASGLPEPPAMNLCWNEIKKSHNLRARLEAFSDHSG
KLQPLQEIIDLWSQDELSAQLPLQGDVALVQKEKTHAFMEEVKSRAPYIYSL
ESAOAFLSQHPFELEPHSESKDTSKQRIQNLRFVWKOATVASELWEKLTARCV
QHRHIERLEQLLEIQGAMEELSTLSQAEGVRATWEPIGDLFIDSLPEHIQAIKLFK
EEFSPMKDGVKLVNDLAHQLAISDVHLSMENSQALEQINRWKQLQASVDERLKQLOD
AHRDFGPGSQHFLSSSVQVPWERAI SPNKVPYVINHQAQCTCWDHPKMTELYQTADL
NNIKFSAYRTAMKLRVQKALRLDLVLTTLTALEIFNEHDLQASEHVMDDVVEVHCLTA
LYERLEEEERGLVNPVLCVDMSLNLLNVFDSGRSGKMRALSEFKTGIAACLCGTEVKEK
LOYLFSOVANSQCQDQRHLGVLLHEAIQVPRQLGEVAAGGNSVPSVRSRFRSTG
KPVIEASQFLEWNLQPMVWLPVLRHTVIAEQVKHTKCSICROCPKIGFRYRSK
QFNVDICQTCFLTGRASKGNKLHYPIMEYPTTSSENMRDFATLLKNKFRSKHYFSK
HPQGYLPVQSVLEADYSETPASSPMWPHADTHSIEHFASRLAEMESQNSFFNDLSL
SPDSDIDEQYLLRHSSPITDREPAFGQAQPCSVATESKGELOKILAHLEDENRILQG
ELRRLKWOHEEAAPSLADGSTAATDHRNEELLAEARILRQHKSRLETRMQILEDH
NKQLESQRLRELLLOPPTESDGSAGSSSLASSPQSGSHPREKQGTTPDTEAAD
DVGSKSDVSLCLEIDIMEKLRHAFPSVRSDDVTANTLLAS"
3193. ->3499
/gene="DRP2"
3'UTR
BASE COUNT 856 a 964 c 913 g 766 t
ORIGIN
Query Match 19.0%; Score 523.4; DB 9; Length 3499;

Db	1572	GGACCTGGTAACCTTAACACAGCCCTGGAAATCTTAATGAGCATGATCTGCAGGCCAG	1631
QY	2503	TGACCCAGCCCATGATATCCTGCAGATTATTAATGTTTGACCACTATTTATGACCCGCT	2562
Db	1632	TGAGCACGTGATGGATGTGGTAGAGTCAATTCACCTGCTGACTGCTTATATGAACGTTT	1691
QY	2563	GGAGCAAGAGCACAAATTTGGTCAACGTCCTCTCTGCGTGGATATGTGTCTGAACGTG	2622
Db	1692	GGAGGAGAAAGAGGCATCCTGGTCAATGTGCCACTCTGTGTGGACATGAGCCTCAATTG	1751
QY	2623	GCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCGTGTCTGTCTTTTAAAC	2682
Db	1752	GCTCCTCAATGTTTGTGATAGTGTGGCAGCGGAAAGATGCGGGCATTTGTCTTTAAGAC	1811
QY	2683	TGGCATCATTTCCCTGTGTAAAGCACATTTTGGAAAGACAAAGTACAGATACCTTTTCAAGCA	2742
Db	1812	TGGCATTCATGCTTGTGTGGCACGGAGTGAAGGAAAAAATTCAGTACCTCTTCAGCCA	1871
QY	2743	AGTGGCAAGTTCA	2755
Db	1872	AGTGGCCAACTCA	1884
RESULT 13			
HSDMDf1			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
COMMENT			
FEATURES			
source			
gene			
CDS			
BASE COUNT			
ORIGIN			
Query Match			
Best Local Similarity			
Matches			
QY	680	AGTAAGCTGATTTGGAACAGGAAATTTATCAGAAGATGAAGAACTGAAGTACAGAGCAG	739
Db	2	AGTAAGCTGATTTGGAACAGGAAATTTATCAGAAGATGAAGAACTGAAGTACAGAGCAG	61

USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
Contact: MGC help desk
Email: cgaps-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbiology.org>
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 46 Row: j Column: 3
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 5032296.

FEATURES
source

1..4658
/organism="Homo sapiens"
/db_xref="LocusID:1756"
/db_xref="taxon:9606"
/clone="MGC:33083 IMAGE:4822807"
/tissue_type="Testis"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Vector: pBluescript"
130..2037
/codon_start=1
/product="dystrophin (muscular dystrophy, Duchenne and
Becker types)"
/protein_id="AAH28720.1"
/db_xref="GI:20379676"
/translation="MREQLKGHTQTTCWDHPKMTELYQSLADLNNVRFSAVRTAMKIL
RRLQKALCLDLLSLAACDALDQHNKQNDQPMIDLIQIINCLTTIYDRLEQEHNNLVN
VPLCVDMCLNLLNVYDTGTRIVLSEFKTGIIISCKAHLEDKYRLFQVASSSTGF
CDQRRLLGLLHDSIQIPRLGEVASFGGSNIPEPSVRSFCQFANNKPEIEAALFLDWMR
LEPQSMVWLPVLRVAAAEATAKHQAKCNICKCEPIIGFRYRSLKFHNYDICQSCFFSG
RVAKGHKMHYPMVEYCTPTTSGEDVRDFAVLKKNKFKTKRYFAKHPRMGYLPVQTVLE
GDNMETPVTLINFWPVDVSAPASSPQLSHDDTHSRIEHYASRLAEMENSGSYLNDNIS
PNESIDDEHLLIQHYQSLNQDSPLSQSPAQILISLESEERGELEIRILADLEEENR
NLQAEYDRLKQHEHKGLSPLSPPEMPTSPQSPRDAELIAEAKLLRQHKGRLEARM
QILEDHNKQLESQHLRLQLLEQPOAEAKVNGTTVSSPSTLQSDSSQPMLLRVVGS
QTSDSMGEEDLLSPPPQDTSTGLEEVMEQLNNSFPSSRGHNVGSLFHMADDLGRAMESL
VSVMTDEEGAE"

CDS

BASE COUNT	1406 a	955 c	934 g	1363 t
ORIGIN				
Query Match	16.6%; Score 457; DB 9; Length 4658;			
Best Local Similarity	100.0%; Pred. No. 5.5e-107;			
Matches 457; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;
QY	2305	CCACGAGACTCAACAACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCGTCTTT	2364	
Db	150	CCACGAGACTCAACAACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCGTCTTT	209	
QY	2365	AGCTGACCTGAATATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	2424	
Db	210	AGCTGACCTGAATATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	269	
QY	2425	GCAGAAGGCCCTTTGCTGGGATCTCTTGAGCCTGTGAGCTGCATGTGATGCCCTTGACCA	2484	
Db	270	GCAGAAGGCCCTTTGCTGGGATCTCTTGAGCCTGTGAGCTGCATGTGATGCCCTTGACCA	329	
QY	2485	GCACAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTGAC	2544	
Db	330	GCACAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTGAC	389	
QY	2545	CACATATTATGACCCGCTGGAGCAAGAGCACAAATTTGGTCAACGTCCTCTCTGCGT	2604	
Db	390	CACATATTATGACCCGCTGGAGCAAGAGCACAAATTTGGTCAACGTCCTCTCTGCGT	449	

QY	2605	GGATATGTGTCTGAACTGGCTGGCTGAATGTTTATGATACGGGACGAGGAGGATCCG	2664
Db	450	GGATATGTGTCTGAACTGGCTGGCTGAATGTTTATGATACGGGACGAGGAGGATCCG	509
QY	2665	TGTCCTGTCTTTTAAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTA	2724
Db	510	TGTCCTGTCTTTTAAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTA	569
QY	2725	CAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGA	2761
Db	570	CAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGA	606

Search completed: April 25, 2003, 14:27:47
Job time : 4786.39 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 07:25:57 ; Search time 384.993 Seconds
(without alignments)
16150.348 Million cell updates/sec

Title: US-09-845-416-9_COPY_540_3300
Perfect score: 2761
Sequence: 1 gccagacctatttgactgga.....aagtggcaagttcaacagga 2761

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4313688

Minimum DB seq length: 0
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*

1:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2761	100.0	3858	24	AAD37237 Human dystrophin m
2	2761	100.0	4825	24	AAD37257 Adeno-associated v
3	2761	100.0	4848	24	AAD37263 Adeno-associated v
4	2610	94.5	3999	24	AAD37234 Human dystrophin m
5	2610	94.5	4966	24	AAD37256 Adeno-associated v
6	2610	94.5	4990	24	AAD37262 Adeno-associated v
7	2427	87.9	4182	24	AAD37230 Human dystrophin m
8	2097	76.0	3531	24	AAD37238 Human dystrophin m
9	2097	76.0	4498	24	AAD37258 Adeno-associated v

10	2055	74.4	3510	24	AAD37240 Human dystrophin m
11	2055	74.4	4476	24	AAD37259 Adeno-associated v
12	1623	58.8	2169	24	AAD37232 Human dystrophin r
13	1623	58.8	4414	24	AAD37260 Adeno-associated v
14	1611	58.3	3446	24	AAD37242 Human dystrophin m
15	1281	46.4	4402	21	AAD37241 A rod shortened dy
16	1276	46.2	1821	24	AAD37241 Human dystrophin r
17	1168.6	42.3	4402	21	AAD37241 A rod shortened dy
18	1152.2	41.7	1991	24	AAD37231 Human dystrophin N
19	1137	41.2	1667	24	AAD37235 Human dystrophin N
20	1078.8	39.1	4075	21	AAD37235 A rod shortened dy
21	889	32.2	1434	24	AAD37239 Human dystrophin r
22	810	29.3	1340	24	AAD37239 Human dystrophin N
23	769.2	27.9	3747	21	AAD37239 A rod shortened dy
24	633.2	22.9	3275	10	AAN97129 Human dystrophin N
25	565.6	20.5	3163	21	AAZ48566 Partial sequence o
26	451	16.3	3172	21	AAZ48570 A rod shortened dy
27	387	14.0	387	24	ABK81990 A rod shortened dy
28	348	12.6	348	24	ABK81989 Human dystrophin s
29	345	12.5	887	24	ABK81993 Human dystrophin s
30	331	12.0	333	24	ABK81967 Human dystrophin C
31	327	11.8	327	24	ABK81966 Human dystrophin s
32	324	11.7	324	24	ABK81991 Human dystrophin s
33	322.2	11.7	702	23	AAS92053 DNA encoding novel
34	255	9.2	255	24	ABK81965 Human dystrophin h
35	226	8.2	756	24	ABK81964 Human dystrophin N
36	215	7.8	215	24	ABK81992 Human dystrophin h
37	198.2	7.2	1500	22	AAF84672 Nucleotide sequenc
38	190.2	6.9	256	22	ABA69736 Human foetal liver
39	190.2	6.9	256	22	ABA36636 Probe #15102 for g
40	190.2	6.9	256	22	AAK17925 Human brain expres
41	190.2	6.9	256	22	AAK43799 Human bone marrow
42	190.2	6.9	256	22	AAI24574 Probe #14507 for g
43	190.2	6.9	256	22	AAI49820 Probe #18506 used
44	190.2	6.9	256	24	ABS18027 Human genome-deriv
45	178	6.4	466	22	ABA58628 Human foetal liver

ALIGNMENTS

RESULT 1
AAD37237
ID AAD37237 standard; DNA; 3858 BP.
XX
AC AAD37237;
DT 21-AUG-2002 (first entry)
XX
DE Human dystrophin minigene delta3849.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Homo sapiens.
XX
PN WO200183695-A2.
PD 08-NOV-2001.
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
XX
PI Xiao X;
XX
DR WPI; 2002-049342/06.
XX
PT New dystrophin minigene for treating Duchenne or Becker muscular dystrophy comprises an N-terminal domain or modified N-terminal domain,

rod repeats, H1 and H4 domains and a cysteine rich domain of a dystrophin gene -

Example 1; Page 48-49; 71pp; English.

The present invention relates to an isolated nucleotide sequence encoding a dystrophin minigene. The minigene comprises N-terminal or modified N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control element. The dystrophin minigene in operable linkage with an expression control element, in a recombinant adeno-associated virus (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is human dystrophin minigene delta3849 containing nucleotides 1-1668 (N-terminus, hinge H1 and rods R1, R2), 8059-10227 (rods R22, R23 and R24, hinge H4 and CR domain) and 11047-11058 (dystrophin last 3 amino acids).

Sequence 3858 BP; 1189 A; 866 C; 905 G; 898 T; 0 other;

Query Match		100.0%;	Score 2761;	DB 24;	Length 3858;
Best Local Similarity		100.0%;	Pred. No. 0;		
Matches 2761; Conservative		0;	Mismatches	0;	Indels
				0;	Gaps
					0;
QY	1	GCCAGACCTATTGACTGGAATAGTGTGGTTGGCTTCCAGCAGTCAGCCACACACGACTGGA	60		
Db	540	GCCAGACCTATTGACTGGAATAGTGTGGTTGGCTTCCAGCAGTCAGCCACACACGACTGGA	599		
QY	61	ACATGCATTCAACATGCCAGATATCAATTAGGCATAGAGAACTACTCGATCCTGGAAGA	120		
Db	600	ACATGCATTCAACATGCCAGATATCAATTAGGCATAGAGAACTACTCGATCCTGGAAGA	659		
QY	121	TGTTGATACCACCTATCCAGATAGAAAGTCCATCTTAAATGTATACATCACATCACTCTTCCA	180		
Db	660	TGTTGATACCACCTATCCAGATAGAAAGTCCATCTTAAATGTATACATCACATCACTCTTCCA	719		
QY	181	AGTTTGGCTCAACAAGTGAGCATTTGAAGCCATCCAGGAAGTGGAAATGTTGCCAAGGCC	240		
Db	720	AGTTTGGCTCAACAAGTGAGCATTTGAAGCCATCCAGGAAGTGGAAATGTTGCCAAGGCC	779		
QY	241	ACCTAAAGTGACTAAAGAAAGAACATTTTCAGTTACATCATCAATGCATATTTCTCAACA	300		
Db	780	ACCTAAAGTGACTAAAGAAAGAACATTTTCAGTTACATCATCAATGCATATTTCTCAACA	839		
QY	301	GATCAGGTCAGTCTAGCACAGGGATATGAGAGAACTTCTCCCTAAGCCTCGATTCAA	360		
Db	840	GATCAGGTCAGTCTAGCACAGGGATATGAGAGAACTTCTCCCTAAGCCTCGATTCAA	899		
QY	361	GAGCTATGCCTACACACAGGCTGCTTATGTACACACCTCTGACCCTACACGGAGCCCAAT	420		
Db	900	GAGCTATGCCTACACACAGGCTGCTTATGTACACACCTCTGACCCTACACGGAGCCCAAT	959		
QY	421	TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAAGTCAATTGGCAGTTTCAATGGAGAG	480		
Db	960	TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAAGTCAATTGGCAGTTTCAATGGAGAG	1019		
QY	481	TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAGTATATATCGTGGCTTCTTC	540		
Db	1020	TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAGTATATATCGTGGCTTCTTC	1079		
QY	541	TGCTGAGGACACATTGCAAGCACACAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA	600		
Db	1080	TGCTGAGGACACATTGCAAGCACACAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA	1139		
QY	601	CCAGTTTCACTACTGAGGGGTACATGATGGATTTGACAGCCCATCAGGCGCGGTTGG	660		
Db	1140	CCAGTTTCACTACTGAGGGGTACATGATGGATTTGACAGCCCATCAGGCGCGGTTGG	1199		
QY	661	TAATATTCTACAAATTGGGAAGTAAGCTGATTGGAACAGGAAATATATCAGAAGATGAAGA	720		
Db	1200	TAATATTCTACAAATTGGGAAGTAAGCTGATTGGAACAGGAAATATATCAGAAGATGAAGA	1259		

QY	721	AACTGAAGTACAAGACAGATGAATCTCTAAATTCAGATGGGAATGCCTCAGGGTAGC	780
Db	1260	AACTGAAGTACAAGACAGATGAATCTCTAAATTCAGATGGGAATGCCTCAGGGTAGC	1319
QY	781	TAGCATGGAACAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAACTCAGAACT	840
Db	1320	TAGCATGGAACAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAACTCAGAACT	1379
QY	841	GAAAGAGTTGAATGACTGGCTAAACAAACAGAAAGAAACAAAGGAAATGGAGGAAGA	900
Db	1380	GAAAGAGTTGAATGACTGGCTAAACAAACAGAAAGAAACAAAGGAAATGGAGGAAGA	1439
QY	901	GCCTCTTGGACCTGATCTTGAAGACCTAAACCCCAAGTACAACAAACATAAGGTGCTTCA	960
Db	1440	GCCTCTTGGACCTGATCTTGAAGACCTAAACCCCAAGTACAACAAACATAAGGTGCTTCA	1499
QY	961	AGAAGATCTAGAACAAACAAAGTCAGGGTCAATTTCTCTCACTCACATGGTGGTAGT	1020
Db	1500	AGAAGATCTAGAACAAACAAAGTCAGGGTCAATTTCTCTCACTCACATGGTGGTAGT	1559
QY	1021	TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGG	1080
Db	1560	TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGG	1619
QY	1081	AGATCGATGGGCAACACATCTGTAGATGGACAGACCCGCTGGGTTCTTTTACAAGACAC	1140
Db	1620	AGATCGATGGGCAACACATCTGTAGATGGACAGACCCGCTGGGTTCTTTTACAAGACAC	1679
QY	1141	TCATAGATTACTGCAACAGTTCCTTACAGGATGCTACAGGATGCTTGGCTGGCTTACAGA	1200
Db	1680	TCATAGATTACTGCAACAGTTCCTTACAGGATGCTTGGCTGGCTTACAGA	1739
QY	1201	AGCTGAAACAACTGCCAATGTCTACAGGATGCTACCCGTAAGGAAAGGCTCCTAGAAGA	1260
Db	1740	AGCTGAAACAACTGCCAATGTCTACAGGATGCTACCCGTAAGGAAAGGCTCCTAGAAGA	1799
QY	1261	CTCCAAGGGAGTAAAGAGCTGATGAAACAATGGCAAGACCTCCAAGGTGAAATTTGAAGC	1320
Db	1800	CTCCAAGGGAGTAAAGAGCTGATGAAACAATGGCAAGACCTCCAAGGTGAAATTTGAAGC	1859
QY	1321	TCACACAGATGTTTATCACAACTGATGAAACAGCCCAAAATTCCTGAGATCCCTGGA	1380
Db	1860	TCACACAGATGTTTATCACAACTGATGAAACAGCCCAAAATTCCTGAGATCCCTGGA	1919
QY	1381	AGGTTCCGATGATGACGCTCTTCAACATTTAGTCCCATTTGGAAGCCAGTTCTGACCAGTG	1440
Db	1920	AGGTTCCGATGATGACGCTCTTCAACATTTAGTCCCATTTGGAAGCCAGTTCTGACCAGTG	1979
QY	1441	TGAACCTCGGAAAAAGTCTCTCAACATTTAGTCCCATTTGGAAGCCAGTTCTGACCAGTG	1500
Db	1980	TGAACCTCGGAAAAAGTCTCTCAACATTTAGTCCCATTTGGAAGCCAGTTCTGACCAGTG	2039
QY	1501	GAAGCGTCTGCACCTTCTCTGCAGGAACTTCTGGTGTGGCTACAGCTGAAAGATGATGA	1560
Db	2040	GAAGCGTCTGCACCTTCTCTGCAGGAACTTCTGGTGTGGCTACAGCTGAAAGATGATGA	2099
QY	1561	ATTAAGCCGGCAGGCACCTATTTGGAGGCGGACTTTCCAGCAGTTTCCAGAGCAGAACGATGT	1620
Db	2100	ATTAAGCCGGCAGGCACCTATTTGGAGGCGGACTTTCCAGCAGTTTCCAGAGCAGAACGATGT	2159
QY	1621	ACATAGGGCCTTCAAGAGGGAATGAAACTAAAGAACCTGTAATCATGAGTACTCTTGA	1680
Db	2160	ACATAGGGCCTTCAAGAGGGAATGAAACTAAAGAACCTGTAATCATGAGTACTCTTGA	2219
QY	1681	GACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAAGGACTAGAGAACTCTACCAGGA	1740
Db	2220	GACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAAGGACTAGAGAACTCTACCAGGA	2279
QY	1741	GCCCAGAGAGCTGCTCCTGAGGAGAGAGCCCAAGTGTCACTCGGCTTCTACGAAGCA	1800
Db	2280	GCCCAGAGAGCTGCTCCTGAGGAGAGAGCCCAAGTGTCACTCGGCTTCTACGAAGCA	2339
QY	1801	GGCTGAGGAGGTCAATACTGAGTGGGAAAAATTTGAACCTGCACCTCCGCTGACTGGCAGAG	1860

Db 2340 GGCTGAGGAGTCAATAGTGGGAAATTTGAACCTGCATCCGCTGACTGGCAGAG 2399
QY 1861 AAAAAATAGATGAGACCTTTGAAAGACTCCAGGAATTCAGAGGCCACGGATGAGCTGGA 1920
Db 2400 AAAAAATAGATGAGACCTTTGAAAGACTCCAGGAATTCAGAGGCCACGGATGAGCTGGA 2459
QY 1921 CCTCAAGCTGCGCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGGATCTCT 1980
Db 2460 CCTCAAGCTGCGCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGGATCTCT 2519
QY 1981 CATTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAATTCGCGC 2040
Db 2520 CATTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAATTCGCGC 2579
QY 2041 TCTGAAAGAGAACCTGAGCCACGTCATGACCTTGCTCGCCAGCTTACCACCTTTGGGCAT 2100
Db 2580 TCTGAAAGAGAACCTGAGCCACGTCATGACCTTGCTCGCCAGCTTACCACCTTTGGGCAT 2639
QY 2101 TCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCTGAACACCCAGATGGAAGCTTCT 2160
Db 2640 TCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCTGAACACCCAGATGGAAGCTTCT 2699
QY 2161 GCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGAAGCCCAACAGGACTTTGGTCC 2220
Db 2700 GCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGAAGCCCAACAGGACTTTGGTCC 2759
QY 2221 AGCATCTCAGCACTTTCTTTCCAGCTCTGTCCAGGCTCCCTGGGAGAGACCCATCTCGCC 2280
Db 2760 AGCATCTCAGCACTTTCTTTCCAGCTCTGTCCAGGCTCCCTGGGAGAGACCCATCTCGCC 2819
QY 2281 AAACAAAGTGCCCTACTATATCAACCCAGGACTCAACCAACTTGCTGGGACCATCCCAA 2340
Db 2820 AAACAAAGTGCCCTACTATATCAACCCAGGACTCAACCAACTTGCTGGGACCATCCCAA 2879
QY 2341 AATGACAGAGCTCTACAGCTTTTAGCTGACCTGAATATGTCAGATTTCTCAGCTTATAG 2400
Db 2880 AATGACAGAGCTCTACAGCTTTTAGCTGACCTGAATATGTCAGATTTCTCAGCTTATAG 2939
QY 2401 GACTGCCATGAACTCCGAAGACTGCAGAGGCGCCCTTTGCTTGGATCTCTTGAGCCCTGTC 2460
Db 2940 GACTGCCATGAACTCCGAAGACTGCAGAGGCGCCCTTTGCTTGGATCTCTTGAGCCCTGTC 2999
QY 2461 AGCTGCATGTGATGCCCTTGGACCAGCACAACCTCAAGCAAAATGACCAGCCCATGGATAT 2520
Db 3000 AGCTGCATGTGATGCCCTTGGACCAGCACAACCTCAAGCAAAATGACCAGCCCATGGATAT 3059
QY 2521 CCTGCAGATTAATTAATTTTGACCACCTATTTATGACCGCCTGGAGCAAGAGCACAACAA 2580
Db 3060 CCTGCAGATTAATTAATTTTGACCACCTATTTATGACCGCCTGGAGCAAGAGCACAACAA 3119
QY 2581 TTTGGTCAACGTCCTCTCTCGGTGGATATGTGTCTGAACCTGGCTGCTGAATGTTATGA 2640
Db 3120 TTTGGTCAACGTCCTCTCTCGGTGGATATGTGTCTGAACCTGGCTGCTGAATGTTATGA 3179
QY 2641 TACGGGACGAACAGGAGGATCCGTGCTCCTGTCTTTTAAACTGGCATCATTTCCCTGTG 2700
Db 3180 TACGGGACGAACAGGAGGATCCGTGCTCCTGTCTTTTAAACTGGCATCATTTCCCTGTG 3239
QY 2701 TAAAGCACATTTGGAAGACAAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGG 2760
Db 3240 TAAAGCACATTTGGAAGACAAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGG 3299
QY 2761 A 2761
Db 3300 A 3300

RESULT 2

AAD37257

ID AAD37257 standard; DNA; 4825 BP.

XX

AC AAD37257;

XX 21-AUG-2002 (first entry)
DT Adeno-associated virus vector plasmid, AAV-MCK-delta3849.
DE Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
XX adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
KW Chimeric - Homo sapiens.
OS Chimeric - Unidentified.
XX WO200183695-A2.
PN 08-NOV-2001.
XX 27-APR-2001; 2001WO-US13677.
PR 28-APR-2000; 2000US-200777P.
XX (XIAO/) XIAO X.
PI Xiao X;
XX WPI; 2002-049342/06.
DR New dystrophin minigene for treating Duchenne or Becker muscular
XX dystrophy comprises an N-terminal domain or modified N-terminal domain,
XX rod repeats, H1 and H4 domains and a cysteine rich domain of a
XX dystrophin gene -
XX Example 1; Page 61-62; 71pp; English.
CC The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is AAV
CC vector plasmid construct containing human dystrophin minigenes, a muscle
CC creatine kinase (MCK) promoter and a small polyA signal sequence.
XX Sequence 4825 BP; 1369 A; 1175 C; 1204 G; 1077 T; 0 other;
SQ
Query Match 100.0%; Score 2761; DB 24; Length 4825;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2761; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCAGACCTATTGACTGGAAATAGTGTGGTTGCCAGCAGTCAGCCACACACGACTGGA 60
Db 1297 GCCAGACCTATTGACTGGAAATAGTGTGGTTGCCAGCAGTCAGCCACACACGACTGGA 1356
QY 61 ACATGCATTCAACATCGCCAGATATCAATAGGCATAGAGAACTACTCGATCCTGAAGA 120
Db 1357 ACATGCATTCAACATCGCCAGATATCAATAGGCATAGAGAACTACTCGATCCTGAAGA 1416
QY 121 TGTGTATACCACTATCCAGATAAAGAGTCCATCTTAATGTACATCATCATCTTCCA 180
Db 1417 TGTGTATACCACTATCCAGATAAAGAGTCCATCTTAATGTACATCATCATCTTCCA 1476
QY 181 AGTTTGCCTCAACAAGTGAAGTGAAGCCATCCAGGAACTGGAATGTTGCCAAGGCC 240
Db 1477 AGTTTGCCTCAACAAGTGAAGTGAAGCCATCCAGGAACTGGAATGTTGCCAAGGCC 1536
QY 241 ACCTAAAGTGACTAAAGAAGAACATTTTCAGTTACATCATCAATGCACATATTCTCAACA 300
Db 1537 ACCTAAAGTGACTAAAGAAGAACATTTTCAGTTACATCATCAATGCACATATTCTCAACA 1596
QY 301 GATCACGGTCACTAGCACAGGGATATGAGAGAACTTCTTCCCCCTAAGCCTCGATTCAA 360

QY 2521 CCTGCAGATTATTAATTGTTGACCACTATTATGACCGCCTGGAGCAAGAGCACAA 2580
Db 3817 CCTGCAGATTATTAATTGTTGACCACTATTATGACCGCCTGGAGCAAGAGCACAA 3876
QY 2581 TTTGGTCAACGTCCTCTCTCGGTGGATATGTGTCTGAACCTGGCTGCTGAATGTTTATGA 2640
Db 3877 TTTGGTCAACGTCCTCTCTCGGTGGATATGTGTCTGAACCTGGCTGCTGAATGTTTATGA 3936
QY 2641 TACGGGACGAACAGGGAGGATCCGTGTCCTCTCTTTTAAACTGGCATCATTTCCCTGTG 2700
Db 3937 TACGGGACGAACAGGGAGGATCCGTGTCCTCTCTTTTAAACTGGCATCATTTCCCTGTG 3996
QY 2701 TAAAGCACATTGGAAGACACAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCACACAGG 2760
Db 3997 TAAAGCACATTGGAAGACACAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCACACAGG 4056
QY 2761 A 2761
Db 4057 A 4057
RESULT 3
AAD37263
ID AAD37263 standard; DNA; 4848 BP.
XX
AC AAD37263;
XX
DT 21-AUG-2002 (first entry)
XX
DE Adeno-associated virus (AAV) vector plasmid, AAV-CMV-delta3849.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Cytomegalovirus.
OS Chimeric - Unidentified.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
XX
PI xiao X;
XX
DR WPI; 2002-049342/06.
XX
PT New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -
XX
PS Example 1; Page 68-70; 71pp; English.
XX
CC The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is AAV
CC vector plasmid construct containing human dystrophin minigenes, a
CC cytomegalovirus (CMV) promoter and a small polyA signal sequence.

XX
SQ Sequence 4848 BP; 1405 A; 1144 C; 1180 G; 1119 T; 0 other;
Query Match
Best Local Similarity 100.0%; Score 2761; DB 24; Length 4848;
Matches 2761; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GCCAGACCTATTGACTGGATAGTGTGGTTGCCAGCAGTCCAGCAACAGACTGGA 60
Db 1320 GCCAGACCTATTGACTGGATAGTGTGGTTGCCAGCAGTCCAGCAACAGACTGGA 1379
QY 61 ACATGCATTCAACATCGCCAGATATCAATTAGGCATAGAGAACTACTCGATCCTGAAGA 120
Db 1380 ACATGCATTCAACATCGCCAGATATCAATTAGGCATAGAGAACTACTCGATCCTGAAGA 1439
QY 121 TGTGATACCCACCTATCCAGATAAGAAAGTCCATCTTAATGTACATCACATCTCTCCA 180
Db 1440 TGTGATACCCACCTATCCAGATAAGAAAGTCCATCTTAATGTACATCACATCTCTCCA 1499
QY 181 AGTTTGGCCTCAACAAAGTGAGCATTGAAGCCATCCAGGAAGTGGAAATGTTGCCAAGGCC 240
Db 1500 AGTTTGGCCTCAACAAAGTGAGCATTGAAGCCATCCAGGAAGTGGAAATGTTGCCAAGGCC 1559
QY 241 ACCTAAAGTGACTAAAGAAGAACATTTTCAGTTACATCATCAATGCACTATTCTCAACA 300
Db 1560 ACCTAAAGTGACTAAAGAAGAACATTTTCAGTTACATCATCAATGCACTATTCTCAACA 1619
QY 301 GATCAGCGTCAGTCTAGCACAGGGATATGAGAGAACTTCTTCCCCTAAGCCTCGATTCAA 360
Db 1620 GATCAGCGTCAGTCTAGCACAGGGATATGAGAGAACTTCTTCCCCTAAGCCTCGATTCAA 1679
QY 361 GAGCTATGCCTACACACAGGCTGCTTATGTACCACCTCTGACCCCTACACGGAGCCCAT 420
Db 1680 GAGCTATGCCTACACACAGGCTGCTTATGTACCACCTCTGACCCCTACACGGAGCCCAT 1739
QY 421 TCCTTCACAGCATTGGAAGCTCTGAAAGACAAAGTCAATTTGGCAGTTCATTGATGGAGAG 480
Db 1740 TCCTTCACAGCATTGGAAGCTCTGAAAGACAAAGTCAATTTGGCAGTTCATTGATGGAGAG 1799
QY 481 TGAAGTAAACCTGGACCGTTATCAACACAGCTTTAGAGAGAGTATTATCGTGGCTTCTTTC 540
Db 1800 TGAAGTAAACCTGGACCGTTATCAACACAGCTTTAGAGAGAGTATTATCGTGGCTTCTTTC 1859
QY 541 TGCTGAGGACACATTTGCAAGCACAAAGGAGAGATTTCTTAATGATGTGGAAGTGGTGAAGA 600
Db 1860 TGCTGAGGACACATTTGCAAGCACAAAGGAGAGATTTCTTAATGATGTGGAAGTGGTGAAGA 1919
QY 601 CCAGTTTCATACTCATGAGGGGTACATGATGGATCTCCTAAATTCAGATGGGAATGCCCTCAGGGTAGC 780
Db 1920 CCAGTTTCATACTCATGAGGGGTACATGATGGATCTCCTAAATTCAGATGGGAATGCCCTCAGGGTAGC 2099
QY 661 TAATATTCTACAATTTGGGAAGTAAGCTGATTTGGAACAGGAAAATTTATCAGAAGATGAAGA 720
Db 1980 TAATATTCTACAATTTGGGAAGTAAGCTGATTTGGAACAGGAAAATTTATCAGAAGATGAAGA 2039
QY 721 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGATGGGAATGCCCTCAGGGTAGC 780
Db 2040 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGATGGGAATGCCCTCAGGGTAGC 2099
QY 781 TAGCATGGAAAAACAAGCAATTTACATAGAGTTTTAATGGATCTCAGAAATCAGAAACT 840
Db 2100 TAGCATGGAAAAACAAGCAATTTACATAGAGTTTTAATGGATCTCAGAAATCAGAAACT 2159
QY 841 GAAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGAAAGAAAATGGAGGAAGA 900
Db 2160 GAAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGAAAGAAAATGGAGGAAGA 2219
QY 901 GCCTCTTGGACCTGATCTTGAAGACCTAAACGCCAAGTACAAACATAGGTTCTTCA 960
Db 2220 GCCTCTTGGACCTGATCTTGAAGACCTAAACGCCAAGTACAAACATAGGTTCTTCA 2279
QY 961 AGAAGATCTAGAACAAAGCAAGTCAAGGTCATCTTCTCACTCACATGGTGGTGTAGT 1020
|||||

2280 AGAAGATCTAGAACAAAGAAAGTCAAGGTCAATTTCTCTCACTCACATGGTGGTAGT 2339
1021 TGATGAATCTAGTGGAGATCACGAACTGCTGCTTTGGGAAGAACAACTTAAGGTATGGG 1080
2340 TGATGAATCTAGTGGAGATCACGAACTGCTGCTTTGGGAAGAACAACTTAAGGTATGGG 2399
1081 AGATCGATGGGCAACATCTGTAGATGGACAGAACCCGCTGGTCTTTTACAAGACAC 1140
2400 AGATCGATGGGCAACATCTGTAGATGGACAGAACCCGCTGGTCTTTTACAAGACAC 2459
1141 TCATAGATTACTGCAACAGTTCCCTGGACCTGGAAAAAGTTTCTTGCCCTGCTTACAGA 1200
2460 TCATAGATTACTGCAACAGTTCCCTGGACCTGGAAAAAGTTTCTTGCCCTGCTTACAGA 2519
1201 AGCTGAAACAACTGCCAATGTCTTACAGGATGCTACCCGTAAGGAAAGGCTCCTAGAAGA 1260
2520 AGCTGAAACAACTGCCAATGTCTTACAGGATGCTACCCGTAAGGAAAGGCTCCTAGAAGA 2579
1261 CTCCAAGGGAGTAAAGAGCTGATGAAACAATGGCAAGACCTCCAAGGTGAAATTTGAAGC 1320
2580 CTCCAAGGGAGTAAAGAGCTGATGAAACAATGGCAAGACCTCCAAGGTGAAATTTGAAGC 2639
1321 TCACACAGATGTTTATCACAACTGGATGAAACAGCCAAAAATCCTGAGATCCCTGGA 1380
2640 TCACACAGATGTTTATCACAACTGGATGAAACAGCCAAAAATCCTGAGATCCCTGGA 2699
1381 AGGTTCCGATGATGCAGTCCCTGATCAAAAGACGTTTGGATAACATGAACCTCAAGTGGAG 1440
2700 AGGTTCCGATGATGCAGTCCCTGATCAAAAGACGTTTGGATAACATGAACCTCAAGTGGAG 2759
1441 TGAACCTCGGAAAAAGTCTCTCAACATTTAGGTCCCATTTGGAAGCCAGTTCTGACCAAGT 1500
2760 TGAACCTCGGAAAAAGTCTCTCAACATTTAGGTCCCATTTGGAAGCCAGTTCTGACCAAGT 2819
1501 GAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGTGGCTACAGCTGAAAGATGATGA 1560
2820 GAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGTGGCTACAGCTGAAAGATGATGA 2879
1561 ATTAAGCCGGCAGGCACCTATTGGAGGCGGACTTTCCAGCAGTTTCAGAAAGCAGACGATGT 1620
2880 ATTAAGCCGGCAGGCACCTATTGGAGGCGGACTTTCCAGCAGTTTCAGAAAGCAGACGATGT 2939
1621 ACATAGGGCCCTCAAGAGGGAATTTGAAACCTTAAAGAACTGTAATCATGAGTACTCTTGA 1680
2940 ACATAGGGCCCTCAAGAGGGAATTTGAAACCTTAAAGAACTGTAATCATGAGTACTCTTGA 2999
1681 GACTGTACGAATATTTCTGCAGAGCAGCCCTTTGGAAAGGACTAGAGAAACTCTACCAGGA 1740
3000 GACTGTACGAATATTTCTGCAGAGCAGCCCTTTGGAAAGGACTAGAGAAACTCTACCAGGA 3059
1741 GCCCAGAGAGCTGCCCTCCTGAGGAGAGAGCCCAAGATGTCACTCGGCTTCTACGAAAGCA 1800
3060 GCCCAGAGAGCTGCCCTCCTGAGGAGAGAGCCCAAGATGTCACTCGGCTTCTACGAAAGCA 3119
1801 GCGTGAAGAGGTCAATACTGAGTGGGAAAAATTTGAACCTGCACCTCCGCTGACTGGCAGAG 1860
3120 GCGTGAAGAGGTCAATACTGAGTGGGAAAAATTTGAACCTGCACCTCCGCTGACTGGCAGAG 3179
1861 AAAAATAGATGAGACCCCTTTGAAAGACTCCAGGAACTTCAAGAGGCCCACGGATGAGCTGGA 1920
3180 AAAAATAGATGAGACCCCTTTGAAAGACTCCAGGAACTTCAAGAGGCCCACGGATGAGCTGGA 3239
1921 CCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGCTGGCGGATCTCCT 1980
3240 CCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGCTGGCGGATCTCCT 3299
1981 CATTGACTCTCTCAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAAAATTCGCC 2040
3300 CATTGACTCTCTCAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAAAATTCGCC 3359
2041 TCTGAAAGAGAACGTGAGCCACGTCATGACCTTGTCTCGCCAGCTTACCACCTTTGGGCAAT 2100
3360 TCTGAAAGAGAACGTGAGCCACGTCATGACCTTGTCTCGCCAGCTTACCACCTTTGGGCAAT 3419

2101 TCAGCTCTCACCGTATACCTCAGCACTCTTGAAGACCTGAAACACAGATGGAAGCTTCT 2160
3420 TCAGCTCTCACCGTATACCTCAGCACTCTTGAAGACCTGAAACACAGATGGAAGCTTCT 3479
2161 GCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGAAGCCACAGGGACTTTGGTCC 2220
3480 GCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGAAGCCACAGGGACTTTGGTCC 3539
2221 AGCATCTCAGCACTTTCTTTCCACGTCCTCCAGGGTCCCTGGGAGAGAGCCATCTCGCC 2280
3540 AGCATCTCAGCACTTTCTTTCCACGTCCTCCAGGGTCCCTGGGAGAGAGCCATCTCGCC 3599
2281 AAACAAAGTGCCCTACTATATCAACACAGGAGACTCAAAACAACTTGTGGGACCATCCCAA 2340
3600 AAACAAAGTGCCCTACTATATCAACACAGGAGACTCAAAACAACTTGTGGGACCATCCCAA 3659
2341 AATGACAGAGCTCTACCACTTTTAGCTGACCTGAATAATGTGAGATTCTCAGCTTATAG 2400
3660 AATGACAGAGCTCTACCACTTTTAGCTGACCTGAATAATGTGAGATTCTCAGCTTATAG 3719
2401 GACTGCCATGAACTCCGAAAGACTGCAGAAAGGCCCTTTGCTGGATCTCTTGAGCCTGTC 2460
3720 GACTGCCATGAACTCCGAAAGACTGCAGAAAGGCCCTTTGCTGGATCTCTTGAGCCTGTC 3779
2461 AGCTGCATGTATGCTTGGACCAAGCAACCTCAAGCAAAATGACCAAGCCCATGGATAT 2520
3780 AGCTGCATGTATGCTTGGACCAAGCAACCTCAAGCAAAATGACCAAGCCCATGGATAT 3839
2521 CCTGCAGATTATTAATGTTTGACCACTATTTATGACCCGCTGGAGCAAGAGCAACAA 2580
3840 CCTGCAGATTATTAATGTTTGACCACTATTTATGACCCGCTGGAGCAAGAGCAACAA 3899
2581 TTTGGTCAACGTCCTCTCTGCTGGATATGTGTGAACCTGGCTGCTGAATGTTTATGA 2640
3900 TTTGGTCAACGTCCTCTCTGCTGGATATGTGTGAACCTGGCTGCTGAATGTTTATGA 3959
2641 TACGGGACCAACAGGGAGGATCCGTCCTGCTTAAACCTGGCATCATTTCCCTGTG 2700
3960 TACGGGACCAACAGGGAGGATCCGTCCTGCTTAAACCTGGCATCATTTCCCTGTG 4019
2701 TAAAGCACATTTGGAAAGACAAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTCAACAG 2760
4020 TAAAGCACATTTGGAAAGACAAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTCAACAG 4079
2761 A 2761
4080 A 4080
RESULT 4
AAD37234
ID AAD37234 standard; DNA; 3999 BP.
XX
AC AAD37234;
XX
DT 21-AUG-2002 (first entry)
XX
DE Human dystrophin minigene delta3990.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Homo sapiens.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.

XX (XIAO/) XIAO X.
XX Xiao X;
XX WPI; 2002-049342/06.
PT New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -
XX
PS Example 1; Page 46-47; 71pp; English.
XX
CC The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is human
CC dystrophin minigene delta3990 containing nucleotides 1-1668 (N-terminus,
CC hinge H1 and rods R1 and R2), 8059-10227 (rods R22, R23 and R24, hinge
CC H4 and CR domain) and 11047-11058 (dystrophin last 3 amino acids).
XX
SQ Sequence 3999 BP; 1223 A; 907 C; 933 G; 936 T; 0 other;

Query Match 94.5%; Score 2610; DB 24; Length 3999;
Best Local Similarity 95.1%; Pred. No. 0;
Matches 2761; Conservative 0; Mismatches 0; Indels 141; Gaps 1;

QY 1 GCCAGACCTATTGACTGGAATAGTGTGGTTGGCCAGCAGTCAGCCACACAAACGACTGGA 60
DB |||||
540 GCCAGACCTATTGACTGGAATAGTGTGGTTGGCCAGCAGTCAGCCACACAAACGACTGGA 599
QY 61 ACATGCATCAACATCGCCAGATATCAATTAGGCATAGAGAACTACTCGATCCTGAAGA 120
DB |||||
600 ACATGCATCAACATCGCCAGATATCAATTAGGCATAGAGAACTACTCGATCCTGAAGA 659
QY 121 TGTGATACCACTATCCAGATAAGAAAGTCCATCTTAATGTACATCACATCACTCTTCCA 180
DB |||||
660 TGTGATACCACTATCCAGATAAGAAAGTCCATCTTAATGTACATCACATCACTCTTCCA 719
QY 181 AGTTTGCCTCAACAAGTGAGCATTTGAAGCCATCCAGGAAGTGGAAATGTTGCCAAGGCC 240
DB |||||
720 AGTTTGCCTCAACAAGTGAGCATTTGAAGCCATCCAGGAAGTGGAAATGTTGCCAAGGCC 779
QY 241 ACCTAAAGTGACTAAAGAAAGAACATTTTCAGTTACATCATCAATGCACATTTCTCAACA 300
DB |||||
780 ACCTAAAGTGACTAAAGAAAGAACATTTTCAGTTACATCATCAATGCACATTTCTCAACA 839
QY 301 GATCAGGTCAGTCTAGCACAGGGATATGAGAGAACTTCTTCCCTAAGCCTCGATTCAA 360
DB |||||
840 GATCAGGTCAGTCTAGCACAGGGATATGAGAGAACTTCTTCCCTAAGCCTCGATTCAA 899
QY 361 GAGCTATGCCTACACACAGGCTGCTTATGTCCACCAAGTCAATTTGGCAGTTTCATTGGAGAG 420
DB |||||
900 GAGCTATGCCTACACACAGGCTGCTTATGTCCACCAAGTCAATTTGGCAGTTTCATTGGAGAG 480
QY 421 TCCTTACAGCATTTGGAAGCTCCTGAAGCAAGTCAATTTGGCAGTTTCATTGGAGAGAG 1019
DB |||||
960 TCCTTACAGCATTTGGAAGCTCCTGAAGCAAGTCAATTTGGCAGTTTCATTGGAGAGAG 1079
QY 481 TGAAGTAAACCTGGACCGCTTATCAAAACAGCTTTAGAAAGAGTATTATCGTGGCTCTTTTC 540
DB |||||
1020 TGAAGTAAACCTGGACCGCTTATCAAAACAGCTTTAGAAAGAGTATTATCGTGGCTCTTTTC 1079
QY 541 TGCTGAGGACACATTTGCAAGCAGGAGAGATTTCTAATGATGTGGAGTGGTGAAGA 600
DB |||||
1080 TGCTGAGGACACATTTGCAAGCAGGAGAGATTTCTAATGATGTGGAGTGGTGAAGA 1139

QY 601 CCAGTTTTCATCTACTGAGGGGTACATGATGATGATTTGACAGCCCATCAGGGCCGGTTGG 660
DB |||||
1140 CCAGTTTTCATCTACTGAGGGGTACATGATGATGATTTGACAGCCCATCAGGGCCGGTTGG 1199
QY 661 TAATATTCTACAAATGGGAAGTAAGCTGATTTGGAACAGGAAATATATCAGAAGATGAAGA 720
DB |||||
1200 TAATATTCTACAAATGGGAAGTAAGCTGATTTGGAACAGGAAATATATCAGAAGATGAAGA 1259
QY 721 AACTGAAGTACAAAGACAGATGAATCTCTTAAATCAAGATGGGAATGCCTCAGGGTAGC 780
DB |||||
1260 AACTGAAGTACAAAGACAGATGAATCTCTTAAATCAAGATGGGAATGCCTCAGGGTAGC 1319
QY 781 TAGCATGGAACAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAACT 840
DB |||||
1320 TAGCATGGAACAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAACT 1379
QY 841 GAAAGAGTTGAATGACTGGCTTAACAAACAGAAAGAAAGAAATGAGGAAAGA 900
DB |||||
1380 GAAAGAGTTGAATGACTGGCTTAACAAACAGAAAGAAAGAAATGAGGAAAGA 1439
QY 901 GCCTCTTGGACCTGATCTTGAAGACCTAAACAGCCAAAGTACAAACATAGGTGCTTCA 960
DB |||||
1440 GCCTCTTGGACCTGATCTTGAAGACCTAAACAGCCAAAGTACAAACATAGGTGCTTCA 1499
QY 961 AGAAGATCTAGAACAAAGCAAGTCAAGTGGGTCAATTTCTCTCACTCAGATGGTGGTAGT 1020
DB |||||
1500 AGAAGATCTAGAACAAAGCAAGTCAAGTGGGTCAATTTCTCTCACTCAGATGGTGGTAGT 1559
QY 1021 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTTGGAGACAACTTAAGGTATTGGG 1080
DB |||||
1560 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTTGGAGACAACTTAAGGTATTGGG 1619
QY 1081 AGATCGATGGGCAACATCTGTAGATGGACAGAAAGACCGCTGGGTCTTTTACAAGAC 1138
DB |||||
1620 AGATCGATGGGCAACATCTGTAGATGGACAGAAAGACCGCTGGGTCTTTTACAAGACCA 1679
QY 1139 ----- 1138
DB GCCTGACCTAGCTCTGGACTGACCACTATTGGAGCCCTCTCTACTCAGACTGTTACTCT 1739
QY 1139 ----- 1138
DB GGTGACACAACTGTGGTTACTAAGGAAACTGCCATCTCCAAACTAGAAATGCCATCTTC 1799
QY 1139 ----- ACTCATAGATTACTGCAACAGTTTCCCTCCCTGGACCTGGAAAA 1179
DB |||||
1800 CTTGATGTTGGAGGTACCTACTCATATAGATTACTGCAACAGTTTCCCTCCCTGGACCTGGAAAA 1859
QY 1180 GTTCTTGCCTGGCTTACAGAAAGTGAACAACTGCCAAATGTCTCTACAGGATGCTACCCG 1239
DB |||||
1860 GTTCTTGCCTGGCTTACAGAAAGTGAACAACTGCCAAATGTCTCTACAGGATGCTACCCG 1919
QY 1240 TAAGGAAAGGCTCCTAGAAAGTCCAAAGGAGTAAAAGAGCTGATGAAACAACTGGCAAGA 1299
DB |||||
1920 TAAGGAAAGGCTCCTAGAAAGTCCAAAGGAGTAAAAGAGCTGATGAAACAACTGGCAAGA 1979
QY 1300 CCTCCAAGGTGAAATGAAGCTCACACAGATGTTTATCACAACTGGATGAAACAACTGGCA 1359
DB |||||
1980 CCTCCAAGGTGAAATGAAGCTCACACAGATGTTTATCACAACTGGATGAAACAACTGGCA 2039
QY 1360 AAAATCCTGAGATCCCTGGAAGGTTCCGATGATGCGAGTCTCTTACAAAGACGTTTGA 1419
DB |||||
2040 AAAATCCTGAGATCCCTGGAAGGTTCCGATGATGCGAGTCTCTTACAAAGACGTTTGA 2099
QY 1420 TAACATGAACCTCAAGTGGAGTGAACCTTCGGAAGAAAGTCTCTCAACATTAGGTCCATT 1479
DB |||||
2100 TAACATGAACCTCAAGTGGAGTGAACCTTCGGAAGAAAGTCTCTCAACATTAGGTCCATT 2159
QY 1480 GGAAGCCAGTTCTGACCCAGTGAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGTG 1539
DB |||||
2160 GGAAGCCAGTTCTGACCCAGTGAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGTG 2219

1540	GCTACAGCTGAAAGATGATGAATTAAAGCCGGCAGGACCTATTGGAGGCGACTTTCACGC	1599
1541		1599
1542		1599
1543		1599
1544		1599
1545		1599
1546		1599
1547		1599
1548		1599
1549		1599
1550		1599
1551		1599
1552		1599
1553		1599
1554		1599
1555		1599
1556		1599
1557		1599
1558		1599
1559		1599
1560	AGTTCAGAACGACGATGTACATAGGGCCCTTCAAGAGGGAATTGAAACTAAAGAACC	1659
1561		1659
1562		1659
1563		1659
1564		1659
1565		1659
1566		1659
1567		1659
1568		1659
1569		1659
1570		1659
1571		1659
1572		1659
1573		1659
1574		1659
1575		1659
1576		1659
1577		1659
1578		1659
1579		1659
1580		1659
1581		1659
1582		1659
1583		1659
1584		1659
1585		1659
1586		1659
1587		1659
1588		1659
1589		1659
1590		1659
1591		1659
1592		1659
1593		1659
1594		1659
1595		1659
1596		1659
1597		1659
1598		1659
1599		1659
1600	AGTTCAGAACGACGATGTACATAGGGCCCTTCAAGAGGGAATTGAAACTAAAGAACC	1659
1601		1659
1602		1659
1603		1659
1604		1659
1605		1659
1606		1659
1607		1659
1608		1659
1609		1659
1610		1659
1611		1659
1612		1659
1613		1659
1614		1659
1615		1659
1616		1659
1617		1659
1618		1659
1619		1659
1620		1659
1621		1659
1622		1659
1623		1659
1624		1659
1625		1659
1626		1659
1627		1659
1628		1659
1629		1659
1630		1659
1631		1659
1632		1659
1633		1659
1634		1659
1635		1659
1636		1659
1637		1659
1638		1659
1639		1659
1640		1659
1641		1659
1642		1659
1643		1659
1644		1659
1645		1659
1646		1659
1647		1659
1648		1659
1649		1659
1650		1659
1651		1659
1652		1659
1653		1659
1654		1659
1655		1659
1656		1659
1657		1659
1658		1659
1659		1659
1660	AGTTCAGAACGACGATGTACATAGGGCCCTTCAAGAGGGAATTGAAACTAAAGAACC	1659
1661		1659
1662		1659
1663		1659
1664		1659
1665		1659
1666		1659
1667		1659
1668		1659
1669		1659
1670		1659
1671		1659
1672		1659
1673		1659
1674		1659
1675		1659
1676		1659
1677		

Db	3300	CTGGCTGCTGAATGTTATGATACGGGACGACAGGAGGATCCGTCTCTTTTAA	3359
QY	2680	AACCTGGCATCATTTCCCTGCTGTAAAGCACATTTGGAAGACAAAGTACAGATACCTTTTCAA	2739
Db	3360	AACCTGGCATCATTTCCCTGCTGTAAAGCACATTTGGAAGACAAAGTACAGATACCTTTTCAA	3419
QY	2740	GCAAGTGGCAAGTTCAACAGGA	2761
Db	3420	GCAAGTGGCAAGTTCAACAGGA	3441
RESULT 5			
AAD37256			
ID	AAD37256 standard; DNA; 4966 BP.		
XX			
AC	AAD37256;		
XX			
DT	21-AUG-2002 (first entry)		
XX	Adeno-associated virus vector plasmid, AAV-MCK-delta3990.		
DE			
XX	Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;		
KW	adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;		
KW	Becker muscular dystrophy; ds.		
XX			
OS	Chimeric - Homo sapiens.		
OS	Chimeric - Unidentified.		
XX			
PN	WO200183695-A2.		
XX	08-NOV-2001.		
PD			
XX	27-APR-2001; 2001WO-US13677.		
PF			
XX	28-APR-2000; 2000US-200777P.		
PR			
XX	(XIAO/) XIAO X.		
PA			
XX	Xiao X;		
PI			
XX	WPI; 2002-049342/06.		
DR			
XX	New dystrophin minigene for treating Duchenne or Becker muscular		
PT	dystrophy comprises an N-terminal domain or modified N-terminal domain,		
PT	rod repeats, H1 and H4 domains and a cysteine rich domain of a		
PT	dystrophin gene -		
XX			
PS	Example 1; Page 59-60; 71pp; English.		
XX	The present invention relates to an isolated nucleotide sequence encoding		
CC	a dystrophin minigene. The minigene comprises N-terminal or modified		
CC	N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4		
CC	domains and cysteine-rich domains of dystrophin or utrophin genes. The		
CC	invention also relates to a recombinant adeno-associated virus (AAV)		
CC	comprising dystrophin minigene operably linked to an expression control		
CC	element. The dystrophin minigene in operable linkage with an expression		
CC	control element, in a recombinant adeno-associated virus or retrovirus is		
CC	useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular		
CC	dystrophy (BMD) in a mammalian subject. The present sequence is AAV		
CC	vector plasmid construct containing human dystrophin minigenes, a muscle		
CC	creatine kinase (MCK) promoter and a small polyA signal sequence.		
XX	Sequence 4966 BP; 1403 A; 1216 C; 1232 G; 1115 T; 0 other;		
SQ			

		Query Match	94.5%;	Score 2610;	DB 24;	Length 4900,
		Best Local Similarity	95.1%;	Pred: No; 0;		
		Matches 2761;	Conservative	0;	Mismatches	0;
					Indels	141;
					Gaps	1;
QY	1	GCACAGACCTATTTGACTGGAAATAGTGTGGTTTCCCGAGCAGTCAGCCACACAACGACTGGA	60			
b	1297	GCACAGACCTATTTGACTGGAAATAGTGTGGTTTGCCAGCAGTCAGCCACACAACGACTGGA	1356			

Mon Apr 28 09:28:38 2003

1b 3517 CCAGCTTACCACCTTTGGGCATTCAGCTCTCACCCGTATAACCTCAGCACTCTGGAAGACCT 3576
2140 GAACACACAGATGGAAGCTTCTGCAGGTGGCCGTCGAGACCGAGTCAGGCAGCTGCATGA 2199
3577 GAACACACAGATGGAAGCTTCTGCAGGTGGCCGTCGAGACCGAGTCAGGCAGCTGCATGA 3636
2200 AGCCACACAGGACTTTGGTCCAGCATCTCAGCACTTTCTTCCACGTCTGTCCAGGTTCC 2259
3637 AGCCACACAGGACTTTGGTCCAGCATCTCAGCACTTTCTTCCACGTCTGTCCAGGTTCC 3696
2260 CTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACACGAGACTCAAAC 2319
3697 CTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACACGAGACTCAAAC 3756
2320 AACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAA 2379
3757 AACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAA 3816
2380 TGTCAGATCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAAGGCTTTG 2439
3817 TGTCAGATCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAAGGCTTTG 3876
2440 CTGGGATCTCTGAGCCTGTGAGCTGCATGTGATGCTGATGCTGATGCTGATGCTGATGCTG 2499
3877 CTGGGATCTCTGAGCCTGTGAGCTGCATGTGATGCTGATGCTGATGCTGATGCTGATGCTG 3936
2500 AAATGACACAGCCCATGGATATCTCTGCAGATTTAATTTGTTGACCACTATTTATGACCG 2559
3937 AAATGACACAGCCCATGGATATCTCTGCAGATTTAATTTGTTGACCACTATTTATGACCG 3996
2560 CCTGGAGCAAGAGCACAACATTTGGTCAACGTCCCTCTCTGCTGCTGATATGTGTCTGAA 2619
3997 CCTGGAGCAAGAGCACAACATTTGGTCAACGTCCCTCTCTGCTGCTGATATGTGTCTGAA 4056
2620 CTGGCTGCTGAATGTTATGATACGGGACGAACAGGAGGATCCGTGCTCTCTTTAA 2679
4057 CTGGCTGCTGAATGTTATGATACGGGACGAACAGGAGGATCCGTGCTCTCTTTAA 4116
2680 AACTGGCATCATTTCCCTGTGTAAAGCACAATTTGGAAGACAAGTACAGATACCTTTTCAA 2739
4117 AACTGGCATCATTTCCCTGTGTAAAGCACAATTTGGAAGACAAGTACAGATACCTTTTCAA 4176

RESULT 6
AAD37262
ID AAD37262 standard; DNA; 4990 BP.
XX
AC AAD37262;
XX
DT 21-AUG-2002 (first entry)
XX
DE Adeno-associated virus (AAV) vector plasmid, AAV-CMV-delta3990.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Cytomegalovirus.
OS Chimeric - Unidentified.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
XX 27-APR-2001; 2001WO-US13677.
XX
XX 28-APR-2000; 2000US-200777P.
XX
XX

(XIAO/) XIAO X.
Xiao X;
WPI; 2002-049342/06.
New dystrophin minigene for treating Duchenne or Becker muscular dystrophy comprises an N-terminal domain or modified N-terminal domain, rod repeats, H1 and H4 domains and a cysteine rich domain of a dystrophin gene -
Example 1; Page 67-68; 71pp; English.
The present invention relates to an isolated nucleotide sequence encoding a dystrophin minigene. The minigene comprises N-terminal or modified N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control element. The dystrophin minigene in operable linkage with an expression control element, in a recombinant adeno-associated virus or retrovirus is useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is AAV vector plasmid construct containing human dystrophin minigenes, a cytomegalovirus (CMV) promoter and a small polyA signal sequence.
Sequence 4990 BP; 1439 A; 1185 C; 1208 G; 1158 T; 0 other;
Query Match 94.5%; Score 2610; DB 24; Length 4990;
Best Local Similarity 95.1%; Pred. No. 0;
Matches 2761; Conservative 0; Mismatches 0; Indels 141; Gaps 1;
QY 1 GCCAGACCTATTGACTGGATAGTGTGGTTGCCAGCAGTCAGCCACACAGACTGGA 60
Db 1321 GCCAGACCTATTGACTGGATAGTGTGGTTGCCAGCAGTCAGCCACACAGACTGGA 1380
QY 61 ACATGCATTCAACATCGCCAGATATCAATAGGCATAGAGAACTACTCGATCCTGAAGA 120
Db 1381 ACATGCATTCAACATCGCCAGATATCAATAGGCATAGAGAACTACTCGATCCTGAAGA 1440
QY 121 TGTTGATACCACTATCCAGATAAGAGTCCATCTTAATGTACATCATCATCTCTTCCA 180
Db 1441 TGTTGATACCACTATCCAGATAAGAGTCCATCTTAATGTACATCATCATCTCTTCCA 1500
QY 181 AGTTTGGCTCAACAAGTGAGCATTTGAAGCCATCCAGGAAGTGGAAATGTTGCCAAGGCC 240
Db 1501 AGTTTGGCTCAACAAGTGAGCATTTGAAGCCATCCAGGAAGTGGAAATGTTGCCAAGGCC 1560
QY 241 ACCTAAAGTGACTAAAGAAGAACATTTTCAGTTACATCATCAATGCACATTTCTCAACA 300
Db 1561 ACCTAAAGTGACTAAAGAAGAACATTTTCAGTTACATCATCAATGCACATTTCTCAACA 1620
QY 301 GATCAGGTCAGTCTAGCACAGGATATGAGAGAACTTCTCCCTAAGCCTCGATTCAA 360
Db 1621 GATCAGGTCAGTCTAGCACAGGATATGAGAGAACTTCTCCCTAAGCCTCGATTCAA 1680
QY 361 GAGCTATGCCTACACACAGGCTGCTTATGTACCACTCTGACCCCTACACGGAGCCCAT 420
Db 1681 GAGCTATGCCTACACACAGGCTGCTTATGTACCACTCTGACCCCTACACGGAGCCCAT 1740
QY 421 TCCTTCACAGCATTTGGAAGCTCCCTGAAGACAGTATGAGAGAACTTCTCCCTAAGCCTCGATTCAA 480
Db 1741 TCCTTCACAGCATTTGGAAGCTCCCTGAAGACAGTATGAGAGAACTTCTCCCTAAGCCTCGATTCAA 1800
QY 481 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAGAAGTATTATCGTGGCTTCTTTC 540
Db 1801 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAGAAGTATTATCGTGGCTTCTTTC 1860
QY 541 TGCTGAGGACACATTGCAAGCACAAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 600
Db 1861 TGCTGAGGACACATTGCAAGCACAAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 1920
QY 601 CCAGTTTCTACTCATGAGGGGTACATGATGATTTTGACAGCCCATCAGGGCCGGTTGG 660

||||| 1921 CCAGTTTCATACACTCATAGGGGTACATGATGGATTTGACAGCCCATCAGGGCCGGTTGG 1980
QY 661 TAATATTCTACAATTGGGAAGTAAGCTGATTGGAACACAGGAAATTTATCAGAAGATGAAGA 720
Db 1981 TAATATTCTACAATTGGGAAGTAAGCTGATTGGAACACAGGAAATTTATCAGAAGATGAAGA 2040
QY 721 AACTGAAGTACAGAGCAGATGAATCTCCTAAATTCAGATGGGAATGCCTCAGGGTAGC 780
Db 2041 AACTGAAGTACAGAGCAGATGAATCTCCTAAATTCAGATGGGAATGCCTCAGGGTAGC 2100
QY 781 TAGCATGGAATAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT 840
Db 2101 TAGCATGGAATAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT 2160
QY 841 GAAAGAGTTGAATGACTGGCTAACAAACAGAGAAAGAAACAAAGGAAATGGAGGAAGA 2220
Db 2161 GAAAGAGTTGAATGACTGGCTAACAAACAGAGAAAGAAACAAAGGAAATGGAGGAAGA 2280
QY 901 GCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCAAAGTACAAACAAATAGGTGCTTCA 960
Db 2221 GCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCAAAGTACAAACAAATAGGTGCTTCA 2280
QY 961 AGAAGATCTAGAACAAAGAACAGTCAAGGTCAATTTCTCTCACTACATGGTGGTAGT 1020
Db 2281 AGAAGATCTAGAACAAAGAACAGTCAAGGTCAATTTCTCTCACTACATGGTGGTAGT 2340
QY 1021 TGATGAATCTAGTGGAGATCAGCAACTGCTGCTTTGGAAGAACAACTTAAAGTATTGGG 1080
Db 2341 TGATGAATCTAGTGGAGATCAGCAACTGCTGCTTTGGAAGAACAACTTAAAGTATTGGG 2400
QY 1081 AGATCGATGGGCAACACATCTGTAGATGGACAGAGACCGCTGGGTTCTTTTACAAGAC -- 1138
Db 2401 AGATCGATGGGCAACACATCTGTAGATGGACAGAGACCGCTGGGTTCTTTTACAAGACCA 2460
QY 1139 ----- 1138
Db 2461 GCCTGACCTAGCTCCTGGACTGACCACATATTGGAGCCTCTCCTACTCAGACTGTTACTCT 2520
QY 1139 ----- 1138
Db 2521 GGTGACACAACCTGTGGTTACTAAGGAACACTGCCATCTCCAAACTAGAAATGCCATCTTC 2580
QY 1139 ----- ACTCATAGATTACTGCAACAGTTCCCCCTGGACCTGGAAAA 1179
Db 2581 CTTGATGTTGGAGGTACCTACTCATAGATTACTGCAACAGTTCCCCCTGGACCTGGAAAA 2640
QY 1180 GTTCTTGCCTGGCTTACAGAAAGTGAACAACTGCCAATGTCTTACAGGATGCTACCCG 1239
Db 2641 GTTCTTGCCTGGCTTACAGAAAGTGAACAACTGCCAATGTCTTACAGGATGCTACCCG 2700
QY 1240 TAAGGAAAGGCTCCTAGAGACTCCAAGGGAGTAAAAGAGCTGATGAAACAAATGGCAAGA 1299
Db 2701 TAAGGAAAGGCTCCTAGAGACTCCAAGGGAGTAAAAGAGCTGATGAAACAAATGGCAAGA 2760
QY 1300 CCTCCAAGGTGAATGAAGCTCACACAGATGTTTATCACAACCTGGATGAAACAGCCA 1359
Db 2761 CCTCCAAGGTGAATGAAGCTCACACAGATGTTTATCACAACCTGGATGAAACAGCCA 2820
QY 1360 AAAAACTCCTGAGATCCCTGGAAGGTTCCGATGATGCAGTCTCTTACAAAGACGTTTGA 1419
Db 2821 AAAAACTCCTGAGATCCCTGGAAGGTTCCGATGATGCAGTCTCTTACAAAGACGTTTGA 2880
QY 1420 TAACATGAATCAAGTGGAGTGAACCTCGGAAAAAGTCTCTCAACATTTAGTCCCATTT 1479
Db 2881 TAACATGAATCAAGTGGAGTGAACCTCGGAAAAAGTCTCTCAACATTTAGTCCCATTT 2940
QY 1480 GGAAGCCAGTTCTGACCAGTGGGAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGTG 1539
Db 2941 GGAAGCCAGTTCTGACCAGTGGGAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGTG 3000
QY 1540 GCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGACCTATTGGAGGCGACTTTCCAGC 1599
|||||

Db 3001 GCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTGGAGGCGACTTTCCAGC 3060
QY 1600 AGTTCAGAAAGCAGAACGATGTACATAGGCGCTTCAAGAGGGGAATTGAAAACTAAAGAACC 1659
Db 3061 AGTTCAGAAAGCAGAACGATGTACATAGGCGCTTCAAGAGGGGAATTGAAAACTAAAGAACC 3120
QY 1660 TGTAAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGAAGG 1719
Db 3121 TGTAAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGAAGG 3180
QY 1720 ACTAGAGAAACTCTACCAGGAGCCAGAGAGCTGCTCCTTGAGGAGAGAGAGAGAGAGTGT 1779
Db 3181 ACTAGAGAAACTCTACCAGGAGCCAGAGAGCTGCTCCTTGAGGAGAGAGAGAGAGAGTGT 3240
QY 1780 CACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGAAAAATTGAACCT 1839
Db 3241 CACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGAAAAATTGAACCT 3300
QY 1840 GCACCTCCGCTGACTGGCAGAGAAAAATAGATGAGACCTTTGAAAGACTCCAGGAACCTCA 1899
Db 3301 GCACCTCCGCTGACTGGCAGAGAAAAATAGATGAGACCTTTGAAAGACTCCAGGAACCTCA 3360
QY 1900 AGAGGCCACGGATGAGCTGGACCTCAAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATCCTG 1959
Db 3361 AGAGGCCACGGATGAGCTGGACCTCAAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATCCTG 3420
QY 1960 GCAGCCCGTGGCGATCTCCTCATTTGACTCTCTCCTCAAGATCACCTCGAGAAAGTCAAGGC 2019
Db 3421 GCAGCCCGTGGCGATCTCCTCATTTGACTCTCTCCTCAAGATCACCTCGAGAAAGTCAAGGC 3480
QY 2020 ACTTCGAGGAGAAAAATGGCGCTCTGAAAGAGAAACGTGAGCCACGTCAATGACCTTGCTCG 2079
Db 3481 ACTTCGAGGAGAAAAATGGCGCTCTGAAAGAGAAACGTGAGCCACGTCAATGACCTTGCTCG 3540
QY 2080 CCAGCTTACCCTTTGGGCATTCAGCTCTCACCCGTATAACCTCAGCACCTCTGGAAGACCT 2139
Db 3541 CCAGCTTACCCTTTGGGCATTCAGCTCTCACCCGTATAACCTCAGCACCTCTGGAAGACCT 3600
QY 2140 GAACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGA 2199
Db 3601 GAACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGA 3660
QY 2200 AGCCACAGGGACTTTGGTCCAGCATCTCAGCACTTCTTCCACGTCTGTCCAGGTCC 2259
Db 3661 AGCCACAGGGACTTTGGTCCAGCATCTCAGCACTTCTTCCACGTCTGTCCAGGTCC 3720
QY 2260 CTGGAGAGAGCCCATCTCGCCAAACAAAGTGCCTACTATATCAACCACGAGACTCAAAAC 2319
Db 3721 CTGGAGAGAGCCCATCTCGCCAAACAAAGTGCCTACTATATCAACCACGAGACTCAAAAC 3780
QY 2320 AACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAA 2379
Db 3781 AACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAA 3840
QY 2380 TGTGAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAAGGCCCTTTG 2439
Db 3841 TGTGAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAAGGCCCTTTG 3900
QY 2440 CTTGGATCTCTTGAGCCTGTGAGTGCATGTGATGCTTGGACCGAGCTGCAGAAAGGCCCTTTG 3960
Db 3901 CTTGGATCTCTTGAGCCTGTGAGTGCATGTGATGCTTGGACCGAGCTGCAGAAAGGCCCTTTG 3960
QY 2500 AAATGACCAGCCCATGGATATCCTGCAGATTATTAATGTTTGGACCTATTTATGACCG 2559
Db 3961 AAATGACCAGCCCATGGATATCCTGCAGATTATTAATGTTTGGACCTATTTATGACCG 4020
QY 2560 CCTGGAGCAAGAGCAGCAACAATTTGGTCAACGTCCTCTCTGCGTGGATATGTGTCTGAA 2619
Db 4021 CCTGGAGCAAGAGCAGCAACAATTTGGTCAACGTCCTCTCTGCGTGGATATGTGTCTGAA 4080
QY 2620 CTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCGTCCTGCTCTTTTAA 2679
Db 4081 CTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCGTCCTGCTCTTTTAA 4140

Mon Apr 28 09:28:38 2003

2680 AACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAAGTACAGATACCTTTTCAA 2739
|||||
4141 AACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAAGTACAGATACCTTTTCAA 4200
|||||
2740 GCAAGTGGCAAGTTCAACAGGA 2761
|||||
4201 GCAAGTGGCAAGTTCAACAGGA 4222
|||||
RESULT 7
AAD37230 standard; DNA; 4182 BP.
ID
XX
AC
AC
AAD37230;
21-AUG-2002 (first entry)
XX
DE
DE
XX
XX
KW
KW
KW
KW
XX
OS
XX
XX
PN
XX
XX
PD
XX
XX
PF
XX
XX
PR
XX
XX
PA
XX
XX
PI
XX
XX
DR
XX
XX
PT
PT
PT
PT
PT
XX
PS
XX
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
XX
SQ
The present invention relates to an isolated nucleotide sequence encoding a dystrophin minigene. The minigene comprises N-terminal or modified N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control element. The dystrophin minigene in operable linkage with an expression control element, in a recombinant adeno-associated virus or retrovirus is useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is human dystrophin minigene delta4173 containing nucleotides 1-1992 (N-terminus, hinge H1 and rods R1, R2 and R3), 8059-10227 (rods R22, R23 and R24, hinge H4 and CR domain) and 11047-11058 (dystrophin last 3 amino acids).
Sequence 4182 BP; 1309 A; 927 C; 970 G; 976 T; 0 other;
Query Match 87.9%; Score 2427; DB 24; Length 4182;
Best Local Similarity 89.5%; Pred. No. 0;
Matches 2761; Conservative 0; Mismatches 0; Indels 324; Gaps 1;
1 GCCAGACCTATTGACTGGAATAGTGTGGTTTGGCAGCAGTCCAGCCACACAAACGACTGGA 60
|||||
540 GCCAGACCTATTGACTGGAATAGTGTGGTTTGGCAGCAGTCCAGCCACACAAACGACTGGA 599
|||||
61 ACATGCATTAACATCGCCAGATATCAATTAGGCATAGAGAACTACTCGATCCTGAAGA 120
|||||

Db 600 ACATGCATTAACATCGCCAGATATCAATTAGGCATAGAGAACTACTCGATCCTGAAGA 659
QY 121 TGTTGATACCACTATCCAGATAAGAGTCCATCTTAATGTACATCACATCACTCTTCCA 180
|||||
Db 660 TGTTGATACCACTATCCAGATAAGAGTCCATCTTAATGTACATCACATCACTCTTCCA 719
|||||
QY 181 AGTTTGGCTCAACAAGTGAGCATTTGAAGCCATCCAGGAAGTGGAAATGTTGCCAAGGCC 240
|||||
Db 720 AGTTTGGCTCAACAAGTGAGCATTTGAAGCCATCCAGGAAGTGGAAATGTTGCCAAGGCC 779
|||||
QY 241 ACCTAAAGTGACTAAAGAAGAACATTTTCAGTTACATCAATGCACATATTCCTCAACA 300
|||||
Db 780 ACCTAAAGTGACTAAAGAAGAACATTTTCAGTTACATCAATGCACATATTCCTCAACA 839
|||||
QY 301 GATCACGGTCAGTCTAGCACAGGGATATGAGAGAACTTCTCCCTTAAGCCTCGATTCAA 360
|||||
Db 840 GATCACGGTCAGTCTAGCACAGGGATATGAGAGAACTTCTCCCTTAAGCCTCGATTCAA 899
|||||
QY 361 GAGCTATGCCTACACACAGGGCTGCTTATGTACACACCTCTGACCTACACGGAGCCCAT 420
|||||
Db 900 GAGCTATGCCTACACACAGGGCTGCTTATGTACACACCTCTGACCTACACGGAGCCCAT 959
|||||
QY 421 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAAGTCAATTTGGCAGTTTCATTTGAGAG 480
|||||
Db 960 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAAGTCAATTTGGCAGTTTCATTTGAGAG 1019
|||||
QY 481 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAGTATTATCGTGGCTTCTTC 540
|||||
Db 1020 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAGTATTATCGTGGCTTCTTC 1079
|||||
QY 541 TGCTGAGGACACATTTGCAAGCACAAAGGAGAGATTTCTAATGATGTGGAAAGTGGTGAAGA 600
|||||
Db 1080 TGCTGAGGACACATTTGCAAGCACAAAGGAGAGATTTCTAATGATGTGGAAAGTGGTGAAGA 1139
|||||
QY 601 CCAGTTTTCATCTACATGAGGGGTACATGATGGATTGACAGCCCATCAGGGCCGGTTGG 660
|||||
Db 1140 CCAGTTTTCATCTACATGAGGGGTACATGATGGATTGACAGCCCATCAGGGCCGGTTGG 1199
|||||
QY 661 TAATATTCTACAATTTGGGAAGTAAAGCTGATTGGAACAGGAAATATTCAGAAGATGAAGA 720
|||||
Db 1200 TAATATTCTACAATTTGGGAAGTAAAGCTGATTGGAACAGGAAATATTCAGAAGATGAAGA 1259
|||||
QY 721 AACTGAAGTACAAGAGCAGATGAATCTCTAAATTCAGATGGAATGCGCTCAGGGTAGC 780
|||||
Db 1260 AACTGAAGTACAAGAGCAGATGAATCTCTAAATTCAGATGGAATGCGCTCAGGGTAGC 1319
|||||
QY 781 TAGCATGGAACAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGATCAGAAACT 840
|||||
Db 1320 TAGCATGGAACAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGATCAGAAACT 1379
|||||
QY 841 GAAAGATTGAATGACTGGCTAACAAAAACAGAAAGAAACAAAGAAATGGAGGAAGA 900
|||||
Db 1380 GAAAGATTGAATGACTGGCTAACAAAAACAGAAAGAAACAAAGAAATGGAGGAAGA 1439
|||||
QY 901 GCCTCTTGGACCTGATCTTGAAGACCTAAACCGCAAGTACAAACATAGGTGCTTCA 960
|||||
Db 1440 GCCTCTTGGACCTGATCTTGAAGACCTAAACCGCAAGTACAAACATAGGTGCTTCA 1499
|||||
QY 961 AGAAGATCTAGAACAAACAAAGTCAAGGGTCAATTTCTCACTCACATGGTGGTAGT 1020
|||||
Db 1500 AGAAGATCTAGAACAAACAAAGTCAAGGGTCAATTTCTCACTCACATGGTGGTAGT 1559
|||||
QY 1021 TGATGAATCTAGTGGAGATCAGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGG 1080
|||||
Db 1560 TGATGAATCTAGTGGAGATCAGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGG 1619
|||||
QY 1081 AGATCGATGGGCAAAACATCTGTAGATGGACAGAACCGCTGGGTCTTTTACAAGAC-- 1138
|||||
Db 1620 AGATCGATGGGCAAAACATCTGTAGATGGACAGAACCGCTGGGTCTTTTACAAGACAT 1679
|||||
QY 1139 ----- 1138
Db 1680 CCTTCTCAAAATGGCAACGCTTACTAGAACAAAGTGCCTTTTTTAGTGCATGGCTTTCAGA 1739
|||||

QY 1139 ----- 1138
Db 1740 AAAAGAAGATGAGTGAACAAGATTACACAACTGGCTTTAAAGATCAAAATGAAATGTT 1799
QY 1139 ----- 1138
Db 1800 ATCAAGTCTTCAAAAACCTGGCCGTTTAAAGCGGATCTAGAAAAGAAAGCAATCCAT 1859
QY 1139 ----- 1138
Db 1860 GGGCAAACTGTATTCACTCAAAACAAGATCTTTCAACACTGAAGAATAAGTCAGTGAC 1919
QY 1139 ----- 1138
Db 1920 CCAGAAGACGGAAGCATGGCTGGATRACTTTGCCCGGTGTTGGGATAATTAGTCCAAAA 1979
QY 1139 ----- ACTCATAGATTACTGCAACAGTTCCCCCTGGACCTGGA 1176
Db 1980 ACTTGAAAAGAGTACAGCACAGACTCATAGATTACTGCAACAGTTCCCCCTGGACCTGGA 2039
QY 1177 AAAGTTTCTTGCCCTGGCTTACAGAAGCTGAAACAACCTGCAAGAGCTGATGAAACAATGGCA 1296
Db 2040 AAAGTTTCTTGCCCTGGCTTACAGAAGCTGAAACAACCTGCAAGAGCTGATGAAACAATGGCA 2099
QY 1237 CCGTAAGGAAGGCTCCTAGAAAGCTCCAAAGGGAGTAAAGAGCTGATGAAACAATGGCA 1296
Db 2100 CCGTAAGGAAGGCTCCTAGAAAGCTCCAAAGGGAGTAAAGAGCTGATGAAACAATGGCA 2159
QY 1297 AGACCTCCAAGGTGAATTGAAGCTCACACAGATGTTTATCACAACTGGATGAAACACAG 1356
Db 2160 AGACCTCCAAGGTGAATTGAAGCTCACACAGATGTTTATCACAACTGGATGAAACACAG 2219
QY 1357 CCAAAAAATCCTGAGATCCCTGGAAGGTTCCGATGATGCAGTCCCTGTTTACAAAGACGTTT 1416
Db 2220 CCAAAAAATCCTGAGATCCCTGGAAGGTTCCGATGATGCAGTCCCTGTTTACAAAGACGTTT 2279
QY 1417 GGATAACATGAATTCGAAGTGAAGTGAAGTTCGGAAGGAGTCTCTCAACATTAGTCCCA 1476
Db 2280 GGATAACATGAATTCGAAGTGAAGTGAAGTTCGGAAGGAGTCTCTCAACATTAGTCCCA 2339
QY 1477 TTTGGAAGCCAGTCTTGACCAGTGGGAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGT 1536
Db 2340 TTTGGAAGCCAGTCTTGACCAGTGGGAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGT 2399
QY 1537 GTGGCTACAGCTGAAAGATGATGAATTAAGCCGAGGACCTATTTGGAGGCGACTTTCC 1596
Db 2400 GTGGCTACAGCTGAAAGATGATGAATTAAGCCGAGGACCTATTTGGAGGCGACTTTCC 2459
QY 1597 AGCAGTTCAAGAGCAGAACGATGTACATAGGCGCTTCAAGAGGGAATTTGAAACTAAAGA 1656
Db 2460 AGCAGTTCAAGAGCAGAACGATGTACATAGGCGCTTCAAGAGGGAATTTGAAACTAAAGA 2519
QY 1657 ACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGA 1716
Db 2520 ACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGA 2579
QY 1717 AGGACTAGAGAAACTCTACCAGGAGCCCAAGAGAGTGCCTCCTGAGGAGAGGCCAGAA 1776
Db 2580 AGGACTAGAGAAACTCTACCAGGAGCCCAAGAGAGTGCCTCCTGAGGAGAGGCCAGAA 2639
QY 1777 TGTCACTCGGCTTCTACGAAGCAGGCTGAGGAGGTCAATACTAGTGGGAAAAAATTGAA 1836
Db 2640 TGTCACTCGGCTTCTACGAAGCAGGCTGAGGAGGTCAATACTAGTGGGAAAAAATTGAA 2699
QY 1837 CCTGCACTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGACTCCAGGAAC 1896
Db 2700 CCTGCACTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGACTCCAGGAAC 2759
QY 1897 TCAAGAGGCCAGGATGAGCTGGACCTCAAGCTGCGCCAAGCTGAGGTGATCAAGGGATC 1956
Db 2760 TCAAGAGGCCAGGATGAGCTGGACCTCAAGCTGCGCCAAGCTGAGGTGATCAAGGGATC 2819

QY 1957 CTGGCAGCCCGTGGCGGATCTCCTCATTTGACTCTCTCCAGATCACCTCGAGAAAGTCAA 2016
Db 2820 CTGGCAGCCCGTGGCGGATCTCCTCATTTGACTCTCTCCAGATCACCTCGAGAAAGTCAA 2879
QY 2017 GGCACCTCGAGGAGAAAATTGCGCCTCTGAAAGAGAAAGCTGAGCCACGTCATGACCTTGC 2076
Db 2880 GGCACCTCGAGGAGAAAATTGCGCCTCTGAAAGAGAAAGCTGAGCCACGTCATGACCTTGC 2939
QY 2077 TCGCAGCTTACCACCTTTGGGCATTTGAGCTCTCAGCTCTCAGCCGTATAACCTCAGCAGCTGGAAGA 2136
Db 2940 TCGCAGCTTACCACCTTTGGGCATTTGAGCTCTCAGCTCTCAGCCGTATAACCTCAGCAGCTGGAAGA 2999
QY 2137 CCTGAACACCAGATGGAAGCTTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCA 2196
Db 3000 CCTGAACACCAGATGGAAGCTTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCA 3059
QY 2197 TGAAGCCACACAGGGACTTTTGGTCCAGCATCTCAGCAGCTTCTTCCAGCTGTCCAGGG 2256
Db 3060 TGAAGCCACACAGGGACTTTTGGTCCAGCATCTCAGCAGCTTCTTCCAGCTGTCCAGGG 3119
QY 2257 TCCCTGGGAGAGAGCCCATCTCGCCAAAACAAAGTGCCTACTATATCAACCCAGAGACTCA 2316
Db 3120 TCCCTGGGAGAGAGCCCATCTCGCCAAAACAAAGTGCCTACTATATCAACCCAGAGACTCA 3179
QY 2317 AACAACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAA 2376
Db 3180 AACAACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAA 3239
QY 2377 TAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAAGGCCCT 2436
Db 3240 TAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAAGGCCCT 3299
QY 2437 TTGCTTGGATCTCTGAGCCTGTGAGCTGCATGTGATGCCTTGGACCAGCACAACTCAA 2496
Db 3300 TTGCTTGGATCTCTGAGCCTGTGAGCTGCATGTGATGCCTTGGACCAGCACAACTCAA 3359
QY 2497 GCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAAATTTGTTGACCACCTATTATGA 2556
Db 3360 GCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAAATTTGTTGACCACCTATTATGA 3419
QY 2557 CCGCCTGGAGCAAGAGACACAACAATTTGGTCAACGTCCTCTCTGCGTGGATATGTGTCT 2616
Db 3420 CCGCCTGGAGCAAGAGACACAACAATTTGGTCAACGTCCTCTCTGCGTGGATATGTGTCT 3479
QY 2617 GAACCTGCTGCTGAATGTTTATGATACGGGACGAAACAGGGAGGATCCGTGTCTCTTT 2676
Db 3480 GAACCTGCTGCTGAATGTTTATGATACGGGACGAAACAGGGAGGATCCGTGTCTCTTT 3539
QY 2677 TAAAACTGGCATCATTTCCCTGTGTAAAGCAGATTTGGAAGACATTTGGAAGACATACCTTTT 2736
Db 3540 TAAAACTGGCATCATTTCCCTGTGTAAAGCAGATTTGGAAGACATTTGGAAGACATACCTTTT 3599
QY 2737 CAAGCAAGTGGCAAGTTCAACACAGGA 2761
Db 3600 CAAGCAAGTGGCAAGTTCAACACAGGA 3624

RESULT 8
AAD37238

ID AAD37238 standard; DNA; 3531 BP.

XX AAD37238;

AC AAD37238;
XX 21-AUG-2002 (first entry)

DT Human dystrophin minigene delta3531.

DE Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;

XX adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;

XX Becker muscular dystrophy; ds.

OS Homo sapiens.

XX

N WO200183695-A2.
X
D 08-NOV-2001.
X
X 27-APR-2001; 2001WO-US13677.
X
X 28-APR-2000; 2000US-200777P.
X
X (XIAO/) XIAO X.
X
X Xiao X;
X WPI; 2002-049342/06.
X
X New dystrophin minigene for treating Duchenne or Becker muscular
X dystrophy comprises an N-terminal domain or modified N-terminal domain,
X rod repeats, H1 and H4 domains and a cysteine rich domain of a
X dystrophin gene -
X
X Example 1; Page 50-51; 71pp; English.
X
X The present invention relates to an isolated nucleotide sequence encoding
X a dystrophin minigene. The minigene comprises N-terminal or modified
X N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
X domains and cysteine-rich domains of dystrophin or utrophin genes. The
X invention also relates to a recombinant adeno-associated virus (AAV)
X comprising dystrophin minigene operably linked to an expression control
X element. The dystrophin minigene in operable linkage with an expression
X control element, in a recombinant adeno-associated virus or retrovirus is
X useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
X dystrophy (BMD) in a mammalian subject. The present sequence is human
X dystrophin minigene delta3531 containing nucleotides 1-1341 (N-terminus,
X hinge H1 and rod R1), 8059-10227 (rods R22, R23 and R24, hinge H4 and
X CR domain) and 11047-11058 (dystrophin last 3 amino acids).
X
X Sequence 3531 BP; 1071 A; 809 C; 824 G; 827 T; 0 other;
X
X Query Match 76.0%; Score 2097; DB 24; Length 3531;
X Best Local Similarity 88.2%; Pred. No. 0;
X Matches 2434; Conservative 0; Mismatches 0; Indels 327; Gaps 1;
X
X 1 GCCAGACCTATTGACTGGAATAGTGTGGTTTGGCCAGCAGTCAGCCACACACGACTGGA 60
X
X 540 GCCAGACCTATTGACTGGAATAGTGTGGTTTGGCCAGCAGTCAGCCACACACGACTGGA 599
X
X 61 ACATGCATTCAACATCGCCAGATATCAATTAGGCATAGAGAACTACTCGATCCTGAAGA 120
X
X 600 ACATGCATTCAACATCGCCAGATATCAATTAGGCATAGAGAACTACTCGATCCTGAAGA 659
X
X 121 TGTGTATACACCTATCCAGATAAAGAGTCCATCTTAATGTACATCATCTCTTCCA 180
X
X 660 TGTGTATACACCTATCCAGATAAAGAGTCCATCTTAATGTACATCATCTCTTCCA 719
X
X 181 AGTTTTCCTCAACAAGTGAGCATTTGAAGCCATCCAGGAGTGGAAATGTTGCCAAGGCC 240
X
X 720 AGTTTTCCTCAACAAGTGAGCATTTGAAGCCATCCAGGAGTGGAAATGTTGCCAAGGCC 779
X
X 241 ACCTAAAGTGACTAAAGAGAACATTTTCAGTTACATCATCAATGCACTATTCTCAACA 300
X
X 780 ACCTAAAGTGACTAAAGAGAACATTTTCAGTTACATCATCAATGCACTATTCTCAACA 839
X
X 301 GATCAGGTCAGTCTAGCACAGGGATATGAGAGAACTTCTTCCCTTAAGCCTCGATTCAA 360
X
X 840 GATCAGGTCAGTCTAGCACAGGGATATGAGAGAACTTCTTCCCTTAAGCCTCGATTCAA 899
X
X 361 GAGCTATGCCTACACACAGGCTGCTTATGTCACCACTCTGACCCCTACACGAGCCCAT 420
X
X 900 GAGCTATGCCTACACACAGGCTGCTTATGTCACCACTCTGACCCCTACACGAGCCCAT 959
X
X 421 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAGAGTCAATTTGGCAGTTTCATTTGAGGAG 480
X
X 960 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAGAGTCAATTTGGCAGTTTCATTTGAGGAG 1019

QY 481 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAGAAAGTATATATCGTGGCTTCTTTC 540
Db 1020 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAGAAAGTATATATCGTGGCTTCTTTC 1079
QY 541 TGCTGAGGACACATTTGCAAGCACAAAGGAGAGATTCTTAATGATGTGGAGTGGTGAAGA 600
Db 1080 TGCTGAGGACACATTTGCAAGCACAAAGGAGAGATTCTTAATGATGTGGAGTGGTGAAGA 1139
QY 601 CCAGTTTCATACTCATGAGGGGTACATGATGGATTGACAGCCCATCAGGGCCGGTGG 660
Db 1140 CCAGTTTCATACTCATGAGGGGTACATGATGGATTGACAGCCCATCAGGGCCGGTGG 1199
QY 661 TAATATTCTACAATTTGGGAAGTAAGCTGATTGGAACAGGAAAATTTATCAGAAAGATGAAGA 720
Db 1200 TAATATTCTACAATTTGGGAAGTAAGCTGATTGGAACAGGAAAATTTATCAGAAAGATGAAGA 1259
QY 721 AACTGAAGTACAGAGCAGATGAATCTCCTAAATTCAGATGGGAATGCCTCAGGGTAGC 780
Db 1260 AACTGAAGTACAGAGCAGATGAATCTCCTAAATTCAGATGGGAATGCCTCAGGGTAGC 1319
QY 781 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT 840
Db 1320 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT 1350
QY 841 GAAAGAGTTGAATGACTGGCTAAACAAAAACAGAAAGAAACAAAGGAAAATGGAGGAAGA 900
Db 1351 ----- 1350
QY 901 GCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCAAGTACAAACAAATAGGTGCTTCA 960
Db 1351 ----- 1350
QY 961 AGAAGATCTAGAACAAAGAACAAAGTCAAGGTCAGGTCATCTCTCACTCAGATGGTGGTAGT 1020
Db 1351 ----- 1350
QY 1021 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAGAAACAACCTTAAGTATTGGG 1080
Db 1351 ----- 1350
QY 1081 AGATCGATGGCAACATCTGTAGATGGACAGAGACCGTGGTCTCTTTTACAGACAC 1140
Db 1351 -----AC 1352
QY 1141 TCATAGATTACTGCAACAGTTCCCTCCCTGGACCTGGAAAAAGTTTCTTGGCTGGCTTACAGA 1200
Db 1353 TCATAGATTACTGCAACAGTTCCCTCCCTGGACCTGGAAAAAGTTTCTTGGCTGGCTTACAGA 1412
QY 1201 AGCTGAAACAACTGCCAATGTCTACAGGATGCTACCCGTAAGGAAAAGGCTCCTAGAAGA 1260
Db 1413 AGCTGAAACAACTGCCAATGTCTACAGGATGCTACCCGTAAGGAAAAGGCTCCTAGAAGA 1472
QY 1261 CTCCAAGGGAGTAAAGAGCTGATGAAACAATGGCAAGACCTCCAAGGTGAAATTTGAAGC 1320
Db 1473 CTCCAAGGGAGTAAAGAGCTGATGAAACAATGGCAAGACCTCCAAGGTGAAATTTGAAGC 1532
QY 1321 TCACACAGATGTTTATCACAACCTGGATGAAACAGCCCAAAAATCCTGAGATCCCTGGA 1380
Db 1533 TCACACAGATGTTTATCACAACCTGGATGAAACAGCCCAAAAATCCTGAGATCCCTGGA 1592
QY 1381 AGGTCCGATGATGCAGTCTCTCAACATTTAGTCCCATTTTGGAGCCAGTTCTGACCAGTG 1440
Db 1593 AGGTCCGATGATGCAGTCTCTCAACATTTAGTCCCATTTTGGAGCCAGTTCTGACCAGTG 1652
QY 1441 TGAATTCGGAAAAAGTCTCTCAACATTTAGTCCCATTTTGGAGCCAGTTCTGACCAGTG 1500
Db 1653 TGAATTCGGAAAAAGTCTCTCAACATTTAGTCCCATTTTGGAGCCAGTTCTGACCAGTG 1712
QY 1501 GAAGCGTCTGCACCTTTCTCTGCAGGAACCTCTGGTGTGGCTACAGCTGAAAGATGATGA 1560
Db 1713 GAAGCGTCTGCACCTTTCTCTGCAGGAACCTCTGGTGTGGCTACAGCTGAAAGATGATGA 1772
QY 1561 ATTAAGCCGCGCAGGCACCTATTGGAGGCGGACTTTCCAGAGTTCAGAAAGCAGAACCATGT 1620

Db 1773 ATTAAGCCGGCAGGCACCTATTGGAGCGGACTTTCCAGCAGTTTCAAGAGCAGAACGATGT 1832
QY 1621 ACATAGGGCCCTTCAAGAGGGAATTGAAAACCTAAAGAACCTGTAATCATGAGTACTCTTGA 1680
Db 1833 ACATAGGGCCCTTCAAGAGGGAATTGAAAACCTAAAGAACCTGTAATCATGAGTACTCTTGA 1892
QY 1681 GACTGTACGGAATATTCTTGACAGAGCAGCCCTTTGGAAGGACTAGAGAAAACCTCTACCAAGGA 1740
Db 1893 GACTGTACGGAATATTCTTGACAGAGCAGCCCTTTGGAAGGACTAGAGAAAACCTCTACCAAGGA 1952
QY 1741 GCCCAGAGAGCTGCTCTCCTGAGGAGAGAGGCCAGAAATGTCACTCGGCTTCTACGAAAAGCA 1800
Db 1953 GCCCAGAGAGCTGCTCTCCTGAGGAGAGAGGCCAGAAATGTCACTCGGCTTCTACGAAAAGCA 2012
QY 1801 GGCTGAGGAGGTCAATACTAGTGGGAAAATTTGAACCTGCACCTCCGCTGACTGGCAGAG 1860
Db 2013 GGCTGAGGAGGTCAATACTAGTGGGAAAATTTGAACCTGCACCTCCGCTGACTGGCAGAG 2072
QY 1861 AAAAATAGATGAGACCCCTTGAAGACTCCAGGAACCTTCAAGAGGCCAGGATGAGCTGGA 1920
Db 2073 AAAAATAGATGAGACCCCTTGAAGACTCCAGGAACCTTCAAGAGGCCAGGATGAGCTGGA 2132
QY 1921 CCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGGATCTCCT 1980
Db 2133 CCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGGATCTCCT 2192
QY 1981 CATTGACTCTCTCCAAAGATCACCTCGAGAAAGTCAAGGACTTCAAGAGGCCAGGATGAGCTGGA 1920
Db 2193 CATTGACTCTCTCCAAAGATCACCTCGAGAAAGTCAAGGACTTCAAGAGGCCAGGATGAGCTGGA 2132
QY 2041 TCTGAAAGAGAGACGTGAGCCACGTCAATGACCTTGTGCGCCAGCTTACCACTTTTGGGCAAT 2100
Db 2253 TCTGAAAGAGAGACGTGAGCCACGTCAATGACCTTGTGCGCCAGCTTACCACTTTTGGGCAAT 2100
QY 2101 TCAGCTCTCACCGTATACCTCAGCACTCTGGAAGACCTGAACACCCAGATGGAAGCTTCT 2160
Db 2313 TCAGCTCTCACCGTATACCTCAGCACTCTGGAAGACCTGAACACCCAGATGGAAGCTTCT 2372
QY 2161 GCAGGTGGCCGTGAGGACCGAGTCAGGCACTGATGAAGCCACAGCCAGCTTTGGTCC 2220
Db 2373 GCAGGTGGCCGTGAGGACCGAGTCAGGCACTGATGAAGCCACAGCCAGCTTTGGTCC 2432
QY 2221 AGCATCTCAGCACTTTCTTTCCAGCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCC 2280
Db 2433 AGCATCTCAGCACTTTCTTTCCAGCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCC 2492
QY 2281 AAACAAAGTGCCCTACTATATCAACACCGAGACTCAAAACACTTGTGCGGACCATCCCAA 2340
Db 2493 AAACAAAGTGCCCTACTATATCAACACCGAGACTCAAAACACTTGTGCGGACCATCCCAA 2552
QY 2341 AATGACAGAGCTCTACCACTTTAGCTGACCTGAATAATGTCAAGATTTCTCAGCTTATAG 2400
Db 2553 AATGACAGAGCTCTACCACTTTAGCTGACCTGAATAATGTCAAGATTTCTCAGCTTATAG 2612
QY 2401 GACTGCCATGAAACTCCGAAGACTGCAGAAAGCCCTTTGCTTGGATCTCTTGAGCCTGTC 2460
Db 2613 GACTGCCATGAAACTCCGAAGACTGCAGAAAGCCCTTTGCTTGGATCTCTTGAGCCTGTC 2672
QY 2461 AGCTGCATGTGATGCTTGGACCAAGCAACCTCAAGCAAAATGACCCAGCCCATGGATAT 2520
Db 2673 AGCTGCATGTGATGCTTGGACCAAGCAACCTCAAGCAAAATGACCCAGCCCATGGATAT 2732
QY 2521 CCTGCAGATTATTAATTGTTGACCACTATTATGACCGCCTGGAGCAAGAGCACAACAA 2580
Db 2733 CCTGCAGATTATTAATTGTTGACCACTATTATGACCGCCTGGAGCAAGAGCACAACAA 2792
QY 2581 TTTGGTCAACGTCCTCTCTGCGTGGATATGTCTGAAGTGGCTGCTGAATGTTTATGA 2640
Db 2793 TTTGGTCAACGTCCTCTCTGCGTGGATATGTCTGAAGTGGCTGCTGAATGTTTATGA 2852
QY 2641 TACGGGACGAACAGGGAGGATCCGTGCTCTTTTAAACCTGGGCATCATTTCCCTGTG 2700

Db 2853 TAGGGACGAACAGGGAGGATCCGTGTCTCTTTTAAACTGGCATCTTCCCTGTG 2912
QY 2701 TAAAGCACATTTTGAAGACAAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCACAGG 2760
Db 2913 TAAAGCACATTTTGAAGACAAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCACAGG 2972
QY 2761 A 2761
Db 2973 A 2973

RESULT 9
AAD37258
ID AAD37258 standard; DNA; 4498 BP.
XX
AC AAD37258;
XX
DT 21-AUG-2002 (first entry)
XX
DE Adeno-associated virus vector plasmid, AAV-MCK-3531.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
XX Becker muscular dystrophy; ds.
OS Chimeric - Homo sapiens.
OS Chimeric - Unidentified.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
XX
PI Xiao X;
XX
DR WPI; 2002-049342/06.
XX

New dystrophin minigene for treating Duchenne or Becker muscular dystrophy comprises an N-terminal domain or modified N-terminal domain, rod repeats, H1 and H4 domains and a cysteine rich domain of a dystrophin gene -

Example 1; Page 62-63; 71pp; English.

The present invention relates to an isolated nucleotide sequence encoding a dystrophin minigene. The minigene comprises N-terminal or modified N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control element. The dystrophin minigene in operable linkage with an expression control element, in a recombinant adeno-associated virus or retrovirus is useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is AAV vector plasmid construct containing human dystrophin minigenes, a muscle creatine kinase (MCK) promoter and a small polyA signal sequence.

Sequence 4498 BP; 1251 A; 1118 C; 1123 G; 1006 T; 0 other;

Query Match
Best Local Similarity 76.0%; Score 2097; DB 24; Length 4498;
Matches 2434; Conservative 0; Mismatches 0; Indels 327; Gaps 1;

QY 1 GCCAGACCTATTGACTGGAATAGTGTGTTGCCAGCAGTCAGCCACACACGACTGGA 60
Db 1297 GCCAGACCTATTGACTGGAATAGTGTGTTGCCAGCAGTCAGCCACACACGACTGGA 1356
QY 61 ACATGCATTCAACATCGCCAGATATCAATTAGGCATAGAGAACTACTCGATCCTGAAGA 120

Db	2110	TCATAGATTACTGCAACAGTTCCCTCGACCTGGAAAAAGTTTCTTGCCTGGCTTACAGA	2169
QY	1201	AGCTGAAACAACCTGCCAATGTCCTACAGGATGCTACCCGTAAGGAAAGGCTCCTAGAAGA	1260
Db	2170	AGCTGAAACAACCTGCCAATGTCCTACAGGATGCTACCCGTAAGGAAAGGCTCCTAGAAGA	2229
QY	1261	CTCCAAGGGAGTAAAGAGCTGATGAAACAATGGCAAGACCTCCAAGGTGAATTTGAAGC	1320
Db	2230	CTCCAAGGGAGTAAAGAGCTGATGAAACAATGGCAAGACCTCCAAGGTGAATTTGAAGC	2289
QY	1321	TCACACAGATGTTTATCACAACTGATGAAACAACAGCCAAAAATCCTGAGATCCCTGGA	1380
Db	2290	TCACACAGATGTTTATCACAACTGATGAAACAACAGCCAAAAATCCTGAGATCCCTGGA	2349
QY	1381	AGGTTCCGATGATGCAGTCTCTCAACATTTAGGTCCTTGGATAACATGAACCTCAAGTGGAG	1440
Db	2350	AGGTTCCGATGATGCAGTCTCTCAACATTTAGGTCCTTGGATAACATGAACCTCAAGTGGAG	2409
QY	1441	TGAACCTCGGAAAAAGTCTCTCAACATTTAGGTCCTTGGAAAGCCAGTTCTGACCAGTG	1500
Db	2410	TGAACCTCGGAAAAAGTCTCTCAACATTTAGGTCCTTGGAAAGCCAGTTCTGACCAGTG	2469
QY	1501	GAAGCGTCTGCACCTTCTCTGCAGGAACCTTCTGGTGTGCTACAGCTGAAAGATGATGA	1560
Db	2470	GAAGCGTCTGCACCTTCTCTGCAGGAACCTTCTGGTGTGCTACAGCTGAAAGATGATGA	2529
QY	1561	ATTAAAGCCGGCAGGACCTATTGGAGGCGACTTTCAGCAGTTTCCAGAGCAGTCAAGAGCAGAACGATGT	1620
Db	2530	ATTAAAGCCGGCAGGACCTATTGGAGGCGACTTTCAGCAGTTTCCAGAGCAGTCAAGAGCAGAACGATGT	2589
QY	1621	ACATAGGGCCCTTCAAGAGGGGAATTGAAAACTAAAGAACCTGTATATCATGAGTACTCTTGA	1680
Db	2590	ACATAGGGCCCTTCAAGAGGGGAATTGAAAACTAAAGAACCTGTATATCATGAGTACTCTTGA	2649
QY	1681	GACTGTACGAATATTCTGACAGAGCAGCCTTTGGAAGGACTAGAGAAACTCTACCAGGA	1740
Db	2650	GACTGTACGAATATTCTGACAGAGCAGCCTTTGGAAGGACTAGAGAAACTCTACCAGGA	2709
QY	1741	GCCCAGAGAGCTGCCCTCCTGAGGAGAGAGCCCAAGATGTCACTCGGCTTCTACGAAAGCA	1800
Db	2710	GCCCAGAGAGCTGCCCTCCTGAGGAGAGAGCCCAAGATGTCACTCGGCTTCTACGAAAGCA	2769
QY	1801	GGCTGAGGAGGTCAATACTGAGTGGGAAAAATTTGAACCTGCACCTCCGCTGACTGGCAGAG	1860
Db	2770	GGCTGAGGAGGTCAATACTGAGTGGGAAAAATTTGAACCTGCACCTCCGCTGACTGGCAGAG	2829
QY	1861	AAAAATAGATGAGACCCCTTGAAGACTCCAGGAACCTCAAGAGGCCACGGATGAGCTGGA	1920
Db	2830	AAAAATAGATGAGACCCCTTGAAGACTCCAGGAACCTCAAGAGGCCACGGATGAGCTGGA	2889
QY	1921	CCTCAAGCTCGCCCAAGCTGAGGTGATCAAGGGATCCTGGACGCCCTGGCGATCTCCT	1980
Db	2890	CCTCAAGCTCGCCCAAGCTGAGGTGATCAAGGGATCCTGGACGCCCTGGCGATCTCCT	2949
QY	1981	CATTGACTCTCTCCAAGATCACTCGAGAAAGTCAAGGCACCTTCGAGGAGAAATTCGCGC	2040
Db	2950	CATTGACTCTCTCCAAGATCACTCGAGAAAGTCAAGGCACCTTCGAGGAGAAATTCGCGC	3009
QY	2041	TCTGAAAGAGAACGTGAGCCACGTCAATGACCTTGTCTGCCAGCTTACCACCTTGGGCAT	2100
Db	3010	TCTGAAAGAGAACGTGAGCCACGTCAATGACCTTGTCTGCCAGCTTACCACCTTGGGCAT	3069
QY	2101	TCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCTGAACACCAGATGGAAGCTTCT	2160
Db	3070	TCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCTGAACACCAGATGGAAGCTTCT	3129
QY	2161	GCAGGTGCCGTGAGGACCGAGTCAGGACGTGAGGACCTGAAGCCACAGGACTTTGGTCC	2220
Db	3130	GCAGGTGCCGTGAGGACCGAGTCAGGACGTGAGGACCTGAAGCCACAGGACTTTGGTCC	3189
QY	2221	AGCATCTCAGCACTTTCTTCCACGTCTGTCCAGGTCCCTGGGAGAGAGCCATCTCGCC	2280
Db	3190	AGCATCTCAGCACTTTCTTCCACGTCTGTCCAGGTCCCTGGGAGAGAGCCATCTCGCC	3249

QY	2281	AAACAAAGTGCCTTACTATATCAACACGAGACTCAAAACAACCTTGCTGGGACCATCCCAA	2344
Db	3250	AAACAAAGTGCCTTACTATATCAACACGAGACTCAAAACAACCTTGCTGGGACCATCCCAA	3309
QY	2341	AATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAATGTGAGATTCTCAGCTTATAG	2400
Db	3310	AATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAATGTGAGATTCTCAGCTTATAG	3369
QY	2401	GACTGCCATGAAACTCCGAAGACTGCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTC	2460
Db	3370	GACTGCCATGAAACTCCGAAGACTGCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTC	3429
QY	2461	AGCTGCATGTGATGCCCTTGGACCAAGCACAACCTCAAGCAAAATGACCAAGCCCATGGATAT	2520
Db	3430	AGCTGCATGTGATGCCCTTGGACCAAGCACAACCTCAAGCAAAATGACCAAGCCCATGGATAT	3489
QY	2521	CCTGCAGATTATTAAATTGTTTGACCACCTATTATGACCGCCTGGAGCAAGAGCACAA	2580
Db	3490	CCTGCAGATTATTAAATTGTTTGACCACCTATTATGACCGCCTGGAGCAAGAGCACAA	3549
QY	2581	TTTGGTCAACGTCCCTCTCTGCGTGGATATGTGCTGCTCTTAAACCTGGCATCATTCCTGTG	2640
Db	3550	TTTGGTCAACGTCCCTCTCTGCGTGGATATGTGCTGCTCTTAAACCTGGCATCATTCCTGTG	3609
QY	2641	TACGGGACCAACAGGAGGATCCGTGCTCCTGCTTTTAAACCTGGCATCATTCCTGTG	2700
Db	3610	TACGGGACCAACAGGAGGATCCGTGCTCCTGCTTTTAAACCTGGCATCATTCCTGTG	3669
QY	2701	TAAAGCACATTGGAAGACAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTCAACAGG	2760
Db	3670	TAAAGCACATTGGAAGACAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTCAACAGG	3729
QY	2761	A 2761	
Db	3730	A 3730	
RESULT 10			
AAD37240			
ID	AAD37240 standard; DNA; 3510 BP.		
XX	AAD37240;		
AC	21-AUG-2002 (first entry)		
XX	Human dystrophin minigene delta3510.		
DT	Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;		
XX	adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;		
DE	Becker muscular dystrophy; ds.		
XX	Homo sapiens.		
KW	WO200183695-A2.		
PN	08-NOV-2001.		
XX	27-APR-2001; 2001WO-US13677.		
PD	28-APR-2000; 2000US-200777P.		
XX	(XIAO/) XIAO X.		
PF	Xiao X;		
XX	WPI; 2002-049342/06.		

Example 1; Page 51-52; 71pp; English.

The present invention relates to an isolated nucleotide sequence encoding a dystrophin minigene. The minigene comprises N-terminal or modified N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control element. The dystrophin minigene in operable linkage with an expression control element, in a recombinant adeno-associated virus or retrovirus is useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is human dystrophin minigene delta3510 containing nucleotides 1-1668 (N-terminus, hinge H1 and rods R1 and R2), 8407-10227 (rods R23 and R24, hinge H4 and CR domain) and 11047-11058 (dystrophin last 3 amino acids).

Sequence 3510 BP; 1073 A; 787 C; 828 G; 822 T; 0 other;

New dystrophin minigene for treating Duchenne or Becker muscular dystrophy comprises an N-terminal domain or modified N-terminal domain, rod repeats, H1 and H4 domains and a cysteine rich domain of a dystrophin gene -

DE Adeno-associated virus vector plasmid, AAV-MCK-3510.
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Unidentified.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
XX
PI Xiao X;
XX
DR WPI; 2002-049342/06.
XX
PT New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -
XX
PS Example 1; Page 63-65; 71pp; English.
XX
CC The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is AAV
CC vector plasmid construct containing human dystrophin minigenes, a muscle
CC creatine kinase (MCK) promoter and a small polyA signal sequence.
XX
SQ Sequence 4476 BP; 1252 A; 1096 C; 1127 G; 1001 T; 0 other;

Query Match
Best Local Similarity 74.4%; Score 2055; DB 24; Length 4476;
Matches 2413; Conservative 0; Mismatches 0; Indels 348; Gaps 1;

QY 1 GCCAGACCTATTGACTGGAATAGTGTGGTTGCCAGCAGTCAGCCACACACGACTGGA 60
Db 1296 GCCAGACCTATTGACTGGAATAGTGTGGTTGCCAGCAGTCAGCCACACACGACTGGA 1355
QY 61 ACATGCATTCACATCGCCAGATATCAATTAGGCATAGAGAACTACTCGATCCTGAAGA 120
Db 1356 ACATGCATTCACATCGCCAGATATCAATTAGGCATAGAGAACTACTCGATCCTGAAGA 1415
QY 121 TGTGATACCACTATCCAGATAAGAGTCCATCTTAATGTACATCACAATCCTTCCA 180
Db 1416 TGTGATACCACTATCCAGATAAGAGTCCATCTTAATGTACATCACAATCCTTCCA 1475
QY 181 AGTTTGCCTCAACAAGTGAGCATTAAGCCATCCAGGAAGTGGAAATGTTGCCAAGGCC 240
Db 1476 AGTTTGCCTCAACAAGTGAGCATTAAGCCATCCAGGAAGTGGAAATGTTGCCAAGGCC 1535
QY 241 ACCTAAAGTGACTAAAGAGAACATTTTCAGTTACATCATCAATGCATATTTCTCAACA 300
Db 1536 ACCTAAAGTGACTAAAGAGAACATTTTCAGTTACATCATCAATGCATATTTCTCAACA 1595
QY 301 GATCACGGTCAGTCTAGCACAGGGATATGAGAGAACTTCTTCCCTAAGCCTCGATTCAA 360
Db 1596 GATCACGGTCAGTCTAGCACAGGGATATGAGAGAACTTCTTCCCTAAGCCTCGATTCAA 1655

QY 361 GAGCTATGCTTACACACACAGGCTGCTTATGTCAACCTCTGACCCCTACACGGAGCCCAT 420
Db 1656 GAGCTATGCTTACACACACAGGCTGCTTATGTCAACCTCTGACCCCTACACGGAGCCCAT 1715
QY 421 TCCTTCACAGCATTTTGAAGCTCCTGAAGACAAGTCAATTTGGCAGTTTCATTGATGGAGAG 480
Db 1716 TCCTTCACAGCATTTTGAAGCTCCTGAAGACAAGTCAATTTGGCAGTTTCATTGATGGAGAG 1775
QY 481 TGAAGTAAACCTGGACCGTTTATCAAAACAGCTTTAGAAGAAGTATTATCGTGGCTTCTTTC 540
Db 1776 TGAAGTAAACCTGGACCGTTTATCAAAACAGCTTTAGAAGAAGTATTATCGTGGCTTCTTTC 1835
QY 541 TGCTGAGGACACATTTGCAAGCACCAAGGAGAGATTTCTAATGATGTGGAAGTGTGAAAGA 600
Db 1836 TGCTGAGGACACATTTGCAAGCACCAAGGAGAGATTTCTAATGATGTGGAAGTGTGAAAGA 1895
QY 601 CCAGTTTTCATCTACTGAGGGGTACATGATGGATTTGACAGCCCATCAGGCGCGGTTGG 660
Db 1896 CCAGTTTTCATCTACTGAGGGGTACATGATGGATTTGACAGCCCATCAGGCGCGGTTGG 1955
QY 661 TAATATTCTACAATTGGGAAGTAAGCTGATTGGAAACAGGAAATTTATCAGAAGATGAAGA 720
Db 1956 TAATATTCTACAATTGGGAAGTAAGCTGATTGGAAACAGGAAATTTATCAGAAGATGAAGA 2015
QY 721 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGATGGGAATGCCTCAGGGTAGC 780
Db 2016 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGATGGGAATGCCTCAGGGTAGC 2075
QY 781 TAGCATGGAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAAACT 840
Db 2076 TAGCATGGAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAAACT 2135
QY 841 GAAAGAGTTGAATGACTGGCTAACAACAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 900
Db 2136 GAAAGAGTTGAATGACTGGCTAACAACAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2195
QY 901 GCCTCTGGACCTGATCTTGAAGACCTAAACGCAAGTACAACAAACATACAGTGTCTTCA 960
Db 2196 GCCTCTGGACCTGATCTTGAAGACCTAAACGCAAGTACAACAAACATACAGTGTCTTCA 2255
QY 961 AGAAGATCTAGAACAAGAACAAAGTCAAGGTCAATCTCTCACTCACATCGTGGTGTAGT 1020
Db 2256 AGAAGATCTAGAACAAGAACAAAGTCAAGGTCAATCTCTCACTCACATCGTGGTGTAGT 2315
QY 1021 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTTGAAGAAAGAAAGAAAGTATGGG 1080
Db 2316 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTTGAAGAAAGAAAGTATGGG 2375
QY 1081 AGATCGATGGGCAACATCTGTAGATGGACAGAACCCGCTGGGTTCTTTTACAAGACAC 1140
Db 2376 AGATCGATGGGCAACATCTGTAGATGGACAGAACCCGCTGGGTTCTTTTACAAGACAC 2432
QY 1141 TCATAGATTACTGCAACAGTTCCCTCGACCTGGAAAAAGTTTCTTGGCTGGCTTACAGA 1200
Db 2433 TCATAGATTACTGCAACAGTTCCCTCGACCTGGAAAAAGTTTCTTGGCTGGCTTACAGA 1260
QY 1201 AGCTGAAACAACTGCCAATGTCTACAGGATGCTACCCGTAAGGAAAGGCTCCTAGAAGA 1260
Db 2433 AGCTGAAACAACTGCCAATGTCTACAGGATGCTACCCGTAAGGAAAGGCTCCTAGAAGA 1320
QY 1261 CTCCAAGGGAGTAAAGAGCTGATGAAACAATGGCAAGACCTCCAAGGTGAAATTTGAAGC 2432
Db 2433 CTCCAAGGGAGTAAAGAGCTGATGAAACAATGGCAAGACCTCCAAGGTGAAATTTGAAGC 1380
QY 1321 TCACACAGATGTTTATCACAACCTGGATGAAACACAGCCAAAAATCCTGAGATCCCTGGA 2432
Db 2433 TCACACAGATGTTTATCACAACCTGGATGAAACACAGCCAAAAATCCTGAGATCCCTGGA 1440
QY 1381 AGGTTCCGATGATGCAGTCTGTTTACAAGACGCTTTCGATAACATGAAGTCAAGTGGAG 2432
Db 2433 AGGTTCCGATGATGCAGTCTGTTTACAAGACGCTTTCGATAACATGAAGTCAAGTGGAG 1500
QY 1441 TGAAGTTCGGAAAAAGTCTCTCAACATTAGGTCCCATTTGGAAGCCAGTTCTTGACCAGTG 1500

Db 3468 CCTGCAGATTATTAATGTTTGAACCACTATTATGACCGCTGGAGCAAGACACAACAA 3527
QY 2581 TTTGGTCAACGTCCTCTCTGCGTGGATATGTCTGAACTGGCTGCTGAATGTTTATGA 2640
Db 3528 TTTGGTCAACGTCCTCTCTGCGTGGATATGTCTGAACTGGCTGCTGAATGTTTATGA 3587
QY 2641 TACGGGACGAACAGGAGGATCCGTGTCCTCTTTTAAACTGGCATCATTTCCCTGTG 2700
Db 3588 TACGGGACGAACAGGAGGATCCGTGTCCTCTTTTAAACTGGCATCATTTCCCTGTG 3647
QY 2701 TAAAGCACATTGGAAGACAAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGG 2760
Db 3648 TAAAGCACATTGGAAGACAAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGG 3707
QY 2761 A 2761
Db 3708 A 3708

RESULT 12

AAD37232 standard; DNA; 2169 BP.

ID AAD37232

XX AAD37232;

AC AAD37232;

XX 21-AUG-2002 (first entry)

DT Human dystrophin rod, hinge and CR domain regions encoding DNA #1.

XX Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;

KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;

KW Becker muscular dystrophy; ds.

XX Homo sapiens.

OS WO200183695-A2.

XX 08-NOV-2001.

XX 27-APR-2001; 2001WO-US13677.

XX 28-APR-2000; 2000US-200777P.

XX (XIAO/) XIAO X.

XX Xiao X;

XX WPI; 2002-049342/06.

PS Example 1; Page 45-46; 71pp; English.

XX The present invention relates to an isolated nucleotide sequence encoding a dystrophin minigene. The minigene comprises N-terminal or modified N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control element. The dystrophin minigene in operable linkage with an expression control element, in a recombinant adeno-associated virus or retrovirus is useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is human dystrophin DNA fragment encoding rods R22, R23 and R24, hinge H4 and CR domain regions.

XX Sequence 2169 BP; 623 A; 529 C; 524 G; 493 T; 0 other;

SQ

Query Match 58.8%; Score 1623; DB 24; Length 2169;

Best Local Similarity 100.0%; Pred. No. 0;

Db 2433 -----CAGTCTGACCAAGT 2447
QY 1501 GAAGCGTCTGCACCTTCTCTGCAGGAACCTTCTGGTGTGGCTACAGCTGAAAGATGATGA 1560
Db 2448 GAAGCGTCTGCACCTTCTCTGCAGGAACCTTCTGGTGTGGCTACAGCTGAAAGATGATGA 2507
QY 1561 ATTAAGCCGGCAGGCACTATTTGGAGGCGGACTTTCCAGCAGTTTCCAGAGCAGAAACGATGT 1620
Db 2508 ATTAAGCCGGCAGGCACTATTTGGAGGCGGACTTTCCAGCAGTTTCCAGAGCAGAAACGATGT 2567
QY 1621 ACATAGGGCCCTTCAAGAGGGAATTGAAACTAAAGAACCTGTAAATCATGAGTACTCTTGA 1680
Db 2568 ACATAGGGCCCTTCAAGAGGGAATTGAAACTAAAGAACCTGTAAATCATGAGTACTCTTGA 2627
QY 1681 GACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAAGGACTAGAGAAACTCTACCAGGA 1740
Db 2628 GACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAAGGACTAGAGAAACTCTACCAGGA 2687
QY 1741 GCCCAGAGAGCTGCTCTCTGAGGAGAGAGCCCAAGATGTCACTCGGCTTCTACGAAAGCA 1800
Db 2688 GCCCAGAGAGCTGCTCTCTGAGGAGAGAGCCCAAGATGTCACTCGGCTTCTACGAAAGCA 2747
QY 1801 GGCTGAGGAGGTCAATACTACTGAGTGGGAAAATTTGAACCTGCACCTCCGCTGACTGGCAGAG 1860
Db 2748 GGCTGAGGAGGTCAATACTACTGAGTGGGAAAATTTGAACCTGCACCTCCGCTGACTGGCAGAG 2807
QY 1861 AAAAATAGATGAGACCTTTGAAAGACTCCAGGAACCTTCAAGAGGCCACGGATGAGCTGGA 1920
Db 2808 AAAAATAGATGAGACCTTTGAAAGACTCCAGGAACCTTCAAGAGGCCACGGATGAGCTGGA 2867
QY 1921 CCTCAAGCTGCGCCAAAGCTGAGTGAAGGATCCTGGCAGCCCGTGGCGGATCTCCT 1980
Db 2868 CCTCAAGCTGCGCCAAAGCTGAGTGAAGGATCCTGGCAGCCCGTGGCGGATCTCCT 2927
QY 1981 CATTGACTCTCTCCAAGATCACTCCGAGCAAGTCAAGGCACTTCGAGGAGAGAAATTCGCGC 2040
Db 2928 CATTGACTCTCTCCAAGATCACTCCGAGCAAGTCAAGGCACTTCGAGGAGAGAAATTCGCGC 2987
QY 2041 TCTGAAGAGAAAGCTGAGCCACGCTGACCTTGCTCGCCAGCTTACCACCTTTGGGCAT 2100
Db 2988 TCTGAAGAGAAAGCTGAGCCACGCTGACCTTGCTCGCCAGCTTACCACCTTTGGGCAT 3047
QY 2101 TCAGCTCTCACCGTATACCTCAGCACTCTGGAAGACCTGAACACCCAGATGGAAGCTTCT 2160
Db 3048 TCAGCTCTCACCGTATACCTCAGCACTCTGGAAGACCTGAACACCCAGATGGAAGCTTCT 3107
QY 2161 GCAGGTGGCCGTCGAGGACCGGAGTCAGGCAGCTGCATGAAGCCACAGGGACTTTGGTCC 2220
Db 3108 GCAGGTGGCCGTCGAGGACCGGAGTCAGGCAGCTGCATGAAGCCACAGGGACTTTGGTCC 3167
QY 2221 AGCATCTCAGCACTTTCTTCCACGCTGTGCCAGGTCCTTGGGAGAGAGCCATCTCGCC 2280
Db 3168 AGCATCTCAGCACTTTCTTCCACGCTGTGCCAGGTCCTTGGGAGAGAGCCATCTCGCC 3227
QY 2281 AAACAAAGTGCCCTACTATATCAACACGAGACTCAACAACTTGTGGGACCATCCCAA 2340
Db 3228 AAACAAAGTGCCCTACTATATCAACACGAGACTCAACAACTTGTGGGACCATCCCAA 3287
QY 2341 AATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAATGTGATGATTTTACAGCTTATAG 2400
Db 3288 AATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAATGTGATGATTTTACAGCTTATAG 3347
QY 2401 GACTGCCATGAAACTCCGAAGACTGCAGAGGCTTGTCTGGATCTCTTGAGCCTGTC 2460
Db 3348 GACTGCCATGAAACTCCGAAGACTGCAGAGGCTTGTCTGGATCTCTTGAGCCTGTC 3407
QY 2461 AGCTGCATGTGATGCTTGGACCCAGCAACCTCAACAAATGACCCAGCCCATGGATAT 2520
Db 3408 AGCTGCATGTGATGCTTGGACCCAGCAACCTCAACAAATGACCCAGCCCATGGATAT 3467
QY 2521 CCTGCAGATTATTAATGTTTGACCACTATTTATGACCGCTGGAGCAAGAGCACAACAA 2580

Matches 1623; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1139	ACTCATAGATTACTGCAACAGTTCCCTGGACCTGGAAAAGTTTCTTGCCCTGGCTTACA	1198
Db	2	ACTCATAGATTACTGCAACAGTTCCCTGGACCTGGAAAAGTTTCTTGCCCTGGCTTACA	61
QY	1199	GAAGCTGAAACAACACTGCCAATGTCTACAGGATGCTACCCGTAAGGAAAGCTCCTAGAA	1258
Db	62	GAAGCTGAAACAACACTGCCAATGTCTACAGGATGCTACCCGTAAGGAAAGCTCCTAGAA	121
QY	1259	GACTCCAAGGGAGTAAAGAGCTGATGAAACAATGGCAAGACCTCCAAGGTGAAATTGAA	1318
Db	122	GACTCCAAGGGAGTAAAGAGCTGATGAAACAATGGCAAGACCTCCAAGGTGAAATTGAA	181
QY	1319	GCTCACACAGATGTTTATCACAACTGCTGATGAAACAATGGCAAGACCTCCAAGGTGAAATTGAA	1378
Db	182	GCTCACACAGATGTTTATCACAACTGCTGATGAAACAATGGCAAGACCTCCAAGGTGAAATTGAA	241
QY	1379	GAAGTTCCGATGATGCAGTCTCTGACAGGACCTTTCGATGAAACAATGGCAAGACCTCCAAGGTGAAATTGAA	1438
Db	242	GAAGTTCCGATGATGCAGTCTCTGACAGGACCTTTCGATGAAACAATGGCAAGACCTCCAAGGTGAAATTGAA	301
QY	1439	AGTGAACCTCGGAAAAAGTCTCTCAACATTAGTCCCATTTGGAGCCAGTTCTTGACCAG	1498
Db	302	AGTGAACCTCGGAAAAAGTCTCTCAACATTAGTCCCATTTGGAGCCAGTTCTTGACCAG	361
QY	1499	TGGAAGCGTCTGCACCTTTCTCTGACAGGAACTTCTGGTGTGGTACAGCTGAAAGATGAT	1558
Db	362	TGGAAGCGTCTGCACCTTTCTCTGACAGGAACTTCTGGTGTGGTACAGCTGAAAGATGAT	421
QY	1559	GAATTAAGCCGGCAGGACCTTATTTGAGGCGACTTCCAGCAGTTTCCAGAGCAGAAACGAT	1618
Db	422	GAATTAAGCCGGCAGGACCTTATTTGAGGCGACTTCCAGCAGTTTCCAGAGCAGAAACGAT	481
QY	1619	GTACATAGGGCCTTCAAGAGGGAATTTGAAACTAAAGAACCTGTAATCATGAGTACTCTT	1678
Db	482	GTACATAGGGCCTTCAAGAGGGAATTTGAAACTAAAGAACCTGTAATCATGAGTACTCTT	541
QY	1679	GAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGAAGGACTAGAGAACTCTACCCAG	1738
Db	542	GAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGAAGGACTAGAGAACTCTACCCAG	601
QY	1739	GAGCCAGAGAGCTGCCTCTCTGAGGAGAGAGCCCAAGATGTCACTCGGCTTCTACGAAAG	1798
Db	602	GAGCCAGAGAGCTGCCTCTCTGAGGAGAGAGCCCAAGATGTCACTCGGCTTCTACGAAAG	661
QY	1799	CAGGCTGAGGAGGTCAATACTAGTGGGAAAAATTTGAACCTGCACCTCCGCTGACTGGCAG	1858
Db	662	CAGGCTGAGGAGGTCAATACTAGTGGGAAAAATTTGAACCTGCACCTCCGCTGACTGGCAG	721
QY	1859	AGAAAAATAGATGAGACCTTGAAGACTCCAGGAACTTCAAGAGGCGCCAGGATGAGCTG	1918
Db	722	AGAAAAATAGATGAGACCTTGAAGACTCCAGGAACTTCAAGAGGCGCCAGGATGAGCTG	781
QY	1919	GACCTCAAGCTGGCCCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGATCTC	1978
Db	782	GACCTCAAGCTGGCCCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGATCTC	841
QY	1979	CTCATTTGACTCTCTCAAGATCACTTCGAGAAAGTCAAGGCACTTCGAGGAGAAATTGCG	2038
Db	842	CTCATTTGACTCTCTCAAGATCACTTCGAGAAAGTCAAGGCACTTCGAGGAGAAATTGCG	901
QY	2039	CCTCTGAAAGAGAACGTGAGCCACAGTCAATGACCTTGTCTGCCAGCTTACCACCTTGGGC	2098
Db	902	CCTCTGAAAGAGAACGTGAGCCACAGTCAATGACCTTGTCTGCCAGCTTACCACCTTGGGC	961
QY	2099	ATTGAGCTCTACCGTATAACCTCAGCAGCTCTGGAAGACCTTGAACACCAAGATGGAAGCTT	2158
Db	962	ATTGAGCTCTACCGTATAACCTCAGCAGCTCTGGAAGACCTTGAACACCAAGATGGAAGCTT	1021
QY	2159	CTGCAGGTGGCGTTCGAGGACCGAGTCAAGGAGCTGATGAAGCCCAAGGAGCTTGGT	2218
Db	1022	CTGCAGGTGGCGTTCGAGGACCGAGTCAAGGAGCTGATGAAGCCCAAGGAGCTTGGT	1081

QY	2219	CCAGCATCTCAGCACCTTTCTTCCACGTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCG	2278
Db	1082	CCAGCATCTCAGCACCTTTCTTCCACGTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCG	1141
QY	2279	CCAAACAAAGTGCCCTACTATATCAACCACGAGACTCAAAACAATTTGCTGGGACCATCCC	2338
Db	1142	CCAAACAAAGTGCCCTACTATATCAACCACGAGACTCAAAACAATTTGCTGGGACCATCCC	1201
QY	2339	AAAATGACAGAGCTCTTACCAGTCTTTAGCTGACCTGAATAATGTCTCAGCTTAT	2398
Db	1202	AAAATGACAGAGCTCTTACCAGTCTTTAGCTGACCTGAATAATGTCTCAGCTTAT	1261
QY	2399	AGGACTGCCATGAAACTCCGAAGACTGCAGAAAGGCGCTTTGCTTGGATCTCTTGAGCCTG	2458
Db	1262	AGGACTGCCATGAAACTCCGAAGACTGCAGAAAGGCGCTTTGCTTGGATCTCTTGAGCCTG	1321
QY	2459	TCAGCTGCATGTGATGCCCTTGACCCAGCACAACCTCAAGCAAAATGACCCAGCCCATGGAT	2518
Db	1322	TCAGCTGCATGTGATGCCCTTGACCCAGCACAACCTCAAGCAAAATGACCCAGCCCATGGAT	1381
QY	2519	ATCCTGCAGATTATTAATTGTTTGACCACTATTATGACCCGCTGGAGCAAGACACAAC	2578
Db	1382	ATCCTGCAGATTATTAATTGTTTGACCACTATTATGACCCGCTGGAGCAAGACACAAC	1441
QY	2579	AATTTGGTCAACGTCCCTCTCTGCTGGGATATGTCTGAACTGGCTGCTGAATGTTTAT	2638
Db	1442	AATTTGGTCAACGTCCCTCTCTGCTGGGATATGTCTGAACTGGCTGCTGAATGTTTAT	1501
QY	2639	GATACGGGACGAAACAGGAGGATCCGTGTCTCTTTAAACTGGCATCATTTCCCTG	2698
Db	1502	GATACGGGACGAAACAGGAGGATCCGTGTCTCTTTAAACTGGCATCATTTCCCTG	1561
QY	2699	TGTAAGCACATTTGGAAGACAAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTCAACA	2758
Db	1562	TGTAAGCACATTTGGAAGACAAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTCAACA	1621
QY	2759	GGA 2761	
Db	1622	GGA 1624	

RESULT 13
AAD37260

ID AAD37260 standard; DNA; 4414 BP.

XX AAD37260;

DT 21-AUG-2002 (first entry)

XX Adeno-associated virus vector plasmid, AAV-MCK-3447.
DE Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.

OS Chimeric - Homo sapiens.

OS Chimeric - Unidentified.

XX WO200183695-A2.

XX 08-NOV-2001.

PF 27-APR-2001; 2001WO-US13677.

XX 28-APR-2000; 2000US-200777P.

PA (XIAO/) XIAO X.

PI Xiao X;

XX WPI; 2002-049342/06.

XX

PT New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -
XX
PS Example 1; Page 65-66; 71pp; English.
XX
CC The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is AAV
CC vector plasmid construct containing human dystrophin minigenes, a muscle
CC creatine kinase (MCK) promoter and a small polyA signal sequence.

XX Sequence 4414 BP; 1255 A; 1075 C; 1086 G; 998 T; 0 other;

Query Match 58.8%; Score 1623; DB 24; Length 4414;
Best Local Similarity 78.4%; Pred. No. 0;
Matches 2165; Conservative 0; Mismatches 185; Indels 411; Gaps 2;

QY	1	GCCAGACCTATTGACTGGGAATAGTGTGGTTTGCCAGCAGTCAGCCACACACGACTGGA	60
Db	1297	GCCAGACCTATTGACTGGGAATAGTGTGGTTTGCCAGCAGTCAGCCACACACGACTGGA	1356
QY	61	ACATGCATTCAACATCGCCAGATATCAATTAGGCATAGAGAACTACTCGATCTGAAGA	120
Db	1357	ACATGCATTCAACATCGCCAGATATCAATTAGGCATAGAGAACTACTCGATCTGAAGA	1416
QY	121	TGTTGATACCACCTATCCAGATATCAATTAGGCATAGAGAACTACTCGATCTTCCA	180
Db	1417	TGTTGATACCACCTATCCAGATATCAATTAGGCATAGAGAACTACTCGATCTTCCA	1476
QY	181	AGTTTGGCTCAACAGTGAGCATTTGAAGCCATCCAGGAAGTGGAAATGTTGCCAAGGCC	240
Db	1477	AGTTTGGCTCAACAGTGAGCATTTGAAGCCATCCAGGAAGTGGAAATGTTGCCAAGGCC	1536
QY	241	ACCTAAAGTGACTAAAGAAACATTTTCAGTTACATCATCAAAATGCACATATCTCAACA	300
Db	1537	ACCTAAAGTGACTAAAGAAACATTTTCAGTTACATCATCAAAATGCACATATCTCAACA	1596
QY	301	GATCAGGTGAGTCTAGCACAGGGATATGAGAGAACTTCTTCCCTAAGCCTCGATTCAA	360
Db	1597	GATCAGGTGAGTCTAGCACAGGGATATGAGAGAACTTCTTCCCTAAGCCTCGATTCAA	1656
QY	361	GAGCTATGCCCTACACACAGGCTGCTTATGTCAACACCTCTGACCCCTACACGGGCCATT	420
Db	1657	GAGCTATGCCCTACACACAGGCTGCTTATGTCAACACCTCTGACCCCTACACGGGCCATT	1716
QY	421	TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAAGTCATTTGGCAGTTTCATTTGAGAG	480
Db	1717	TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAAGTCATTTGGCAGTTTCATTTGAGAG	1776
QY	481	TGAAGTAAACCTGGACCGGTTATCAAAACAGCTTTAGAAAGAGTATTATCGTGGCTTCTTC	540
Db	1777	TGAAGTAAACCTGGACCGGTTATCAAAACAGCTTTAGAAAGAGTATTATCGTGGCTTCTTC	1836
QY	541	TGCTGAGGACACATTCGAAGCACAAGGAGAGATTCTTAATGATGTGGAAAGTGGTGAAGA	600
Db	1837	TGCTGAGGACACATTCGAAGCACAAGGAGAGATTCTTAATGATGTGGAAAGTGGTGAAGA	1896
QY	601	CCAGTTTCATCTACATGAGGGGTACATGATGGATTGACAGCCCATCAGGGCCGGTTGG	660
Db	1897	CCAGTTTCATCTACATGAGGGGTACATGATGGATTGACAGCCCATCAGGGCCGGTTGG	1956
QY	661	TAATATTCTACAATTTGGGAAGTAAAGCTGATTGGAACAGGAAATATCAGAAGATGAAGA	720
Db	1957	TAATATTCTACAATTTGGGAAGTAAAGCTGATTGGAACAGGAAATATCAGAAGATGAAGA	2016

QY	721	AACCTGAAGTACAAGAGCAGATGAATCTCTAAATTCAGATGGGAATGCCTCAGGGTAGC	780
Db	2017	AACCTGAAGTACAAGAGCAGATGAATCTCTAAATTCAGATGGGAATGCCTCAGGGTAGC	2076
QY	781	TAGCATGGAACAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT	840
Db	2077	TAGCATGGAACAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT	2136
QY	841	GAAAGAGTTGAATGACTGGCTAAACAAACAGAGAAAGAAATGGAGGAAGA	900
Db	2137	GAAAGAGTTGAATGACTGGCTAAACAAACAGAGAAAGAAATGGAGGAAGA	2196
QY	901	GCCTCTTGGACCTGATCTTGAAGACCTAAACGCCCAAGTACAACAACATAAGGTGCTTCA	960
Db	2197	GCCTCTTGGACCTGATCTTGAAGACCTAAACGCCCAAGTACAACAACATAAGGTGCTTCA	2256
QY	961	AGAAGATCTAGAACAAGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT	1020
Db	2257	AGAAGATCTAGAACAAGCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT	2316
QY	1021	TGATGAATCTAGTGGAGATCAGCAACTGCTGCTTTGGGAAGAACAACTTAAGGTATTGGG	1080
Db	2317	TGATGAATCTAGTGGAGATCAGCAACTGCTGCTTTGGGAAGAACAACTTAAGGTATTGGG	2376
QY	1081	AGATCGATGGCAAAACATCTGTAGATGGACAGAACGCCGCTGGTTCTTTTACAAGACAC	1140
Db	2377	AGATCGATGGCAAAACATCTGTAGATGGACAGAACGCCGCTGGTTCTTTTACAAGACAT	2436
QY	1141	TCATAGATTACTGCAACAGTTTCCCTCCCTGGACCTGGAAGAAAGTTTCTTGCCTGGCTTACAGA	1200
Db	2437	CTTCTCAATGGCAACGCTTACTTGAAGAACAGTGCCTTTTGTAGTGCATGGCTTTCAGA	2496
QY	1201	AGCTGAACAACTGCCAATGTCTTACAGGATGCTACCCGTAAGAAAGGCTCCTTACAAGA	1260
Db	2497	AAAAGAGATGCAGTGAACAAGATTTCACACAACTGGCTTTA-----AAGA	2541
QY	1261	CTCCAAGGGAGTAAAGAGCTGATGAACAATGGCAAGACCTCCAAGGTGAAATTTGAAGC	1320
Db	2542	TCAAAATGAAATGTTATCAAGTCTTCAAAAACCTGGCCGTTTAAAGCGGATCTAGAAA	2601
QY	1321	TCACACAGATGTTTATCAACACCTGGATGAAACAGCCAAACAAATCCTGAGATCCCTGGA	1380
Db	2602	GAAAAGCAATCCATGGSCAAACTGTATTCACTCAACAAAGATCTTCTTTCAACACTGAA	2661
QY	1381	AGTTTCCGATGATGCAGTCCCTTACAAAGACGTTTGGATACATGAACATTTCAAGTGGAG	1440
Db	2662	GAATAAGTCAGTGACCCAGAACGGAAGCATGGCTGGATAACTTTGCCCGGTGTTGGGA	2721
QY	1441	TGAACCTCGGAAAGTCTCTCAACATTAGGTCCTTATTTGGAAAGCAGTCTTGACCAGTG	1500
Db	2722	TAATTTAGTCCAAAACCTTGAAAAGAGT-----	2749
QY	1501	GAAAGCTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGGCTACAGCTGAAAGATGATGA	1560
Db	2750	-----	2749
QY	1561	ATTAAGCCGCGAGGCACCTATTGGAGGCGACTTTCCAGCAGTTTCAGAAAGCAGAACGATGT	1620
Db	2750	-----	2749
QY	1621	ACATAGGCGCTTCAAGAGGGAATTGAAAACCTAAAGAACCTGTATCATGAGTACTCTTGA	1680
Db	2750	-----	2749
QY	1681	GACTGTACGAATATTCTGTACAGAGCAGCCTTTGGAGGACTAGAGAACTCTACCAGGA	1740
Db	2750	-----	2749
QY	1741	GCCCAGAGAGCTGCCCTCTGTAGGAGAGAGCCAGAAATGTCACTCGGCTTCTACGAAAGCA	1800
Db	2750	-----	2749

QY 1801 GGCTGAGGAGGTCAATACTGAGTGGGAAAAATTGAACCTGCACTCCGCTGACTGGCAGAG 1860
Db 2750 ----- 2749
QY 1861 AAAATAGATGAGACCCCTTGAAGACTCCAGGAACCTCAAGAGGCCACGGATGAGCTGGA 1920
Db 2750 ----ACAGCACAGACCCCTTGAAGACTCCAGGAACCTCAAGAGGCCACGGATGAGCTGGA 2805
QY 1921 CCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGGATCTCCT 1980
Db 2806 CCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGGATCTCCT 2865
QY 1981 CATTGACTCTCTCCAGATACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAATTCGCGCC 2040
Db 2866 CATTGACTCTCTCCAGATACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAATTCGCGCC 2925
QY 2041 TCTGAAAGAGAACGTCGAGCCACGTCACCTCAATGACCTTGCTCGCCAGCTTACCACCTTTGGGCAT 2100
Db 2926 TCTGAAAGAGAACGTCGAGCCACGTCACCTCAATGACCTTGCTCGCCAGCTTACCACCTTTGGGCAT 2985
QY 2101 TCAGCTCTCACCGTATATACTCAGCACTCTGGAAGACCTGAACACCCAGATGGAAGCTTCT 2160
Db 2986 TCAGCTCTCACCGTATATACTCAGCACTCTGGAAGACCTGAACACCCAGATGGAAGCTTCT 3045
QY 2161 GCAGGTGGCCGTCGAGACCGGAGTCAGGCAGCTGCATGAAGCCACAGGGACTTTGGTCC 2220
Db 3046 GCAGGTGGCCGTCGAGACCGGAGTCAGGCAGCTGCATGAAGCCACAGGGACTTTGGTCC 3105
QY 2221 AGCATCTCAGCACTTTCTTTCCACGCTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCC 2280
Db 3106 AGCATCTCAGCACTTTCTTTCCACGCTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCC 3165
QY 2281 AAACAAAGTGCCCTACTATATCAACCCAGAGACTCAACAACTTGCTGGGACCATCCCAA 2340
Db 3166 AAACAAAGTGCCCTACTATATCAACCCAGAGACTCAACAACTTGCTGGGACCATCCCAA 3225
QY 2341 AATGACAGAGCTCTACCACTTTAGCTGACCTGAATAATGTCTAGATTTCTCAGCTTATAG 2400
Db 3226 AATGACAGAGCTCTACCACTTTAGCTGACCTGAATAATGTCTAGATTTCTCAGCTTATAG 3285
QY 2401 GACTGCCATGAAACTCCGAAGACTGCAGAGGCCCTTTGCTTGGATCTCTTGAGCCTGTC 2460
Db 3286 GACTGCCATGAAACTCCGAAGACTGCAGAGGCCCTTTGCTTGGATCTCTTGAGCCTGTC 3345
QY 2461 AGCTGCATGTGATGCCCTTGGACCAAGCAACCTCAAGCAAAATGACCCAGCCCATGATAT 2520
Db 3346 AGCTGCATGTGATGCCCTTGGACCAAGCAACCTCAAGCAAAATGACCCAGCCCATGATAT 3405
QY 2521 CCTGCAGATTATTAATTGTTGACCACTATTATGACCGCCTGGAGCAAGAGCACAACAA 2580
Db 3406 CCTGCAGATTATTAATTGTTGACCACTATTATGACCGCCTGGAGCAAGAGCACAACAA 3465
QY 2581 TTTGGTCAACGTCCTCTGCGTGGATATGTGTCTGAACCTGGCTGCTGAATGTTTATGA 2640
Db 3466 TTTGGTCAACGTCCTCTGCGTGGATATGTGTCTGAACCTGGCTGCTGAATGTTTATGA 3525
QY 2641 TACGGGACGAACAGGGAGGATCCGCTGCTCTTTTAAACTGGCATCATTTCCCTGTG 2700
Db 3526 TACGGGACGAACAGGGAGGATCCGCTGCTCTTTTAAACTGGCATCATTTCCCTGTG 3585
QY 2701 TAAAGCACATTTGGAAGACAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGG 2760
Db 3586 TAAAGCACATTTGGAAGACAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGG 3645
QY 2761 A 2761
Db 3646 A 3646

RESULT 14

AD37242

D AAD37242 standard; DNA; 3446 BP.

X

AC

XX

DT

XX

DE

XX

KW

KW

KW

XX

OS

XX

PN

XX

PD

XX

PF

XX

PR

XX

PA

XX

PI

XX

DR

XX

PT

PT

PT

PT

PS

XX

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

CC

AAD37242;

21-AUG-2002 (first entry)

Human dystrophin minigene delta3447.

Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
Becker muscular dystrophy; ds.

Homo sapiens.

WO200183695-A2.

08-NOV-2001.

27-APR-2001; 2001WO-US13677.

28-APR-2000; 2000US-200777P.

(XIAO/) XIAO X.

Xiao X;

WPI; 2002-049342/06.

New dystrophin minigene for treating Duchenne or Becker muscular
dystrophy comprises an N-terminal domain or modified N-terminal domain,
rod repeats, H1 and H4 domains and a cysteine rich domain of a
dystrophin gene -

Example 1; Page 53-54; 71pp; English.

The present invention relates to an isolated nucleotide sequence encoding
a dystrophin minigene. The minigene comprises N-terminal or modified
N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
domains and cysteine-rich domains of dystrophin or utrophin genes. The
invention also relates to a recombinant adeno-associated virus (AAV)
comprising dystrophin minigene operably linked to an expression control
element. The dystrophin minigene in operable linkage with an expression
control element, in a recombinant adeno-associated virus or retrovirus is
useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
dystrophy (BMD) in a mammalian subject. The present sequence is human
dystrophin minigene delta3447 containing nucleotides 1-1992 (N-terminus,
hinge H1 and rods R1, R2 and R3), 8749-10227 (rod R24, hinge H4 and
CR domain) and 11047-11058 (dystrophin last 3 amino acids).

Sequence 3446 BP; 1074 A; 766 C; 787 G; 819 T; 0 other;

Query Match

Best Local Similarity 58.3%; Score 1611; DB 24; Length 3446;

Matches 2164; Conservative 0; Mismatches 185; Indels 412; Gaps 3;

QY 1 GCCAGACCTATTGACTGGAATAGTGTGGTTGCCAGCAGTCAGCCACACACGACTGGA 60

Db 540 GCCAGACCTATTGACTGGAATAGTGTGGTTGCCAGCAGTCAGCCACACACGACTGGA 599

QY 61 ACATGCATTCAACATCGCCAGATATCAATTAGGCATAGAGAACTACTCGATCCTGAAGA 120

Db 600 ACATGCATTCAACATCGCCAGATATCAATTAGGCATAGAGAACTACTCGATCCTGAAGA 659

QY 121 TGTGTATACCACTATCCAGATAGAAAGTCCATCTTAATGTACATCACATCTTCCA 180

Db 660 TGTGTATACCACTATCCAGATAGAAAGTCCATCTTAATGTACATCACATCTTCCA 719

QY 181 AGTTTTCCTCAACAAAGTGAGCATTGAAGCCATCCAGGAACTGGAATGTTGCCAAGGCC 240

Db 720 AGTTTTCCTCAACAAAGTGAGCATTGAAGCCATCCAGGAACTGGAATGTTGCCAAGGCC 779

QY 241 ACCTAAAGTGACTAAAGAAAGAACATTTTCAGTTACATCATCAATGCACTATTCTCAACA 300

Db 780 ACCTAAAGTGACTAAAGAAAGAACATTTTCAGTTACATCATCAATGCACTATTCTCAACA 839

Db	1904	GAATAAGT	CAGT	GACCCAGAAAGACGGAAGCATGGCTGGATAACTTTGCCCGGTGTGGGA	1963
QY	1441	TGAAACTTCGGAAAAAGTCTCTCAACATTAGGTCCCATTTTGGAAAGCCAGTTCTGACCAGTG	1500		
Db	1964	TAATTAGTCCAAAAAACTTGAAAAGAGTACAGCAC			1998
QY	1501	GAAGCGTCTCACCTTTCTCTGCAGGAACCTTCTGGTGTGGCTACAGCTGAAAGATGATGA	1560		
Db	1999				1998
QY	1561	ATTAAGCCGCGCAGGCACCTATTGGAGGCGACTTTCAGAGCAGTTCAGAAGCAGAACGATGT	1620		
Db	1999				1998
QY	1621	ACATAGGGCCCTTCAAGAGGGAATTGAAAACTAAAGAAACCTGTAATCATGAGTACTCTTGA	1680		
Db	1999				1998
QY	1681	GACTGTACGAATATTTCTGACAGCAGCAGCCTTTTGAAGGACTAGAGAAACTCTACCAGGA	1740		
Db	1999				1998
QY	1741	GCCCAGAGAGCTGCCTCCTGAGGAGAGAGCCCAAGATGTCACTCGGCTTCTACGAAAAGCA	1800		
Db	1999				1998
QY	1801	GGCTGAGGAGGTCAATACTAGTGGGAAAAATTGAACCTGCACCTCCGCTGACTGGCAGAG	1860		
Db	1999				1998
QY	1861	AAAAATAGATGAGACCCCTTTGAAAGACTCCAGGAACCTTCAAGAGGCCACGGATGAGCTGGA	1920		
Db	1999				2047
QY	1921	CCTCAAGCTCGCCAAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCCGTGGCGGATCTCCT	1980		
Db	2048	CCTCAAGCTCGCCAAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCCGTGGCGGATCTCCT	2107		
QY	1981	CATTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAATTCGCC	2040		
Db	2108	CATTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAATTCGCC	2167		
QY	2041	TCTGAAAGAGAACTGAGCCACGTCACCTCAATGACCTTGCCTCGCCAGCTTACCACCTTTGGGCA	2100		
Db	2168	TCTGAAAGAGAACTGAGCCACGTCACCTCAATGACCTTGCCTCGCCAGCTTACCACCTTTGGGCA	2227		
QY	2101	TCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCTGAACACCAGATGGAAGCTTCT	2160		
Db	2228	TCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCTGAACACCAGATGGAAGCTTCT	2287		
QY	2161	GCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGAAGCCCACAGGGACTTTGGTTC	2220		
Db	2288	GCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGAAGCCCACAGGGACTTTGGTTC	2347		
QY	2221	AGCATCTCAGCACTTTCTTTCCACGTCGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCC	2280		
Db	2348	AGCATCTCAGCACTTTCTTTCCACGTCGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCC	2407		
QY	2281	AAACAAAAGTCCCCTACTATATCAACCACGAGACTCAAAACAACTGTCTGGGACCATCCCAA	2340		
Db	2408	AAACAAAAGTCCCCTACTATATCAACCACGAGACTCAAAACAACTGTCTGGGACCATCCCAA	2467		
QY	2341	AATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAATGTGAGATCTCAGCTTATAG	2400		
Db	2468	AATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAATGTGAGATCTCAGCTTATAG	2527		
QY	2401	GACTGCCATGAAACTCCGAAGACTGCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTC	2460		
Db	2528	GACTGCCATGAAACTCCGAAGACTGCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTC	2587		
QY	2461	AGCTGCATGTGATGCCTTGGACCAGCAACCTCAAGCAAAATGACCCCATGATAT	2520		

Mon Apr 28 09:28:38 2003

Db 1805 ----- 1804
QY 1141 TCATAGATTACTGCAACAGTTCCCCCTGGACCTGGAAAAGTTTCTTGCCCTGGCTTACAGA 1200
Db 1805 ----- 1804
QY 1201 AGCTGAAACAACCTGCCAATGTCTTACAGGATGCTACCCGTAAGGAAAGGCTCCTAGAGA 1260
Db 1805 ----- 1804
QY 1261 CTCCAAGGAGTAAAGAGCTGATGAACAATGGCAAGACCTTCCAAGGTGAATTTGAAGC 1320
Db 1805 ----- 1804
QY 1321 TCACACAGATGTTTATCACAACTGGATGAAAAACAGCCAAAAATCCTGAGATCCCTGGA 1380
Db 1805 ----- 1804
QY 1381 AGGTTCCGATGATGCAGTCCCTGTTACAAAGACGTTTGGATAACATGAACCTCAAGTGGAG 1440
Db 1805 ----- 1804
QY 1441 TGAACCTCGGAAAAAGTCTCTCAACATTAGGTCCCATTTGGAAGCCAGTTCTGACCAGTG 1500
Db 1805 ----- 1804
QY 1501 GAAGCGCTGCACCTTTCTCTGCAGGAACCTTCTGGTGTGCTACAGCTGAAAGATGATGA 1560
Db 1805 ----- 1804
QY 1561 ATTAAGCGCGCAGGCACCTATTGGAGCGGACTTTCCAGCAGTTTCAAGAGCAGAACGATGT 1620
Db 1805 ----- 1804
QY 1621 ACATAGGCGCTTCAAGAGGGAATTGAAAACCTAAAGAACCTGTAAATCATGAGTACTCTTGA 1680
Db 1805 ----- 1804
QY 1681 GACTGTACGAATATTCTGACAGAGCAGCCCTTTGGAAGGACTAGAGAAACTCTACCAGGA 1740
Db 1805 ----- 1804
QY 1741 GCCCAGAGAGCTGCCTCTCTGAGGAGAGAGCCCAAGATGTCACTCGGCTTCTACGAAAGCA 1800
Db 1805 ----- 1804
QY 1801 GGCTGAGGAGGTCAATACTAGTGGGAAAAATTGAACCTGCACCTCCGCTGACTGGCAGAG 1860
Db 1805 ----- AGGAGGTCAATACTAGTGGGAAAAATTGAACCTGCACCTCCGCTGACTGGCAGAG 1859
QY 1861 AAAATATAGATGAGACCCCTTGAAGACTCCAGGAACCTTCAAGAGGCCACGGATGAGCTGGA 1920
Db 1860 AAAATATAGATGAGACCCCTTGAAGACTCCAGGAACCTTCAAGAGGCCACGGATGAGCTGGA 1919
QY 1921 CCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGGATCTCCT 1980
Db 1920 CCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGGATCTCCT 1979
QY 1981 CATTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAAAATTGCGCC 2040
Db 1980 CATTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAAAATTGCGCC 2039
QY 2041 TCTGAAAGAGAACCTGAGCCACAGTCAATGACCTTGTCTGCCAGCTTACCACCTTTGGGCAT 2100
Db 2040 TCTGAAAGAGAACCTGAGCCACAGTCAATGACCTTGTCTGCCAGCTTACCACCTTTGGGCAT 2099
QY 2101 TCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCTTGAACACCAGATGGAAGCTTCT 2160
Db 2100 TCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCTTGAACACCAGATGGAAGCTTCT 2159
QY 2161 GCAGGTGGCCGTCGAGGACCGAGTCAGGACGTGCATGAAGCCACAGGGACTTTGGTCC 2220
Db 2160 GCAGGTGGCCGTCGAGGACCGAGTCAGGACGTGCATGAAGCCACAGGGACTTTGGTCC 2219

QY 2221 AGCATCTCAGACATTTCTTCCACGCTGTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCC 2280
Db 2220 AGCATCTCAGACATTTCTTCCACGCTGTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCC 2279
QY 2281 AAACAAAAGTGCCCTACTATATCAACCACGAGACTCAAAACAACCTTGTGGGACCATCCCAA 2340
Db 2280 AAACAAAAGTGCCCTACTATATCAACCACGAGACTCAAAACAACCTTGTGGGACCATCCCAA 2339
QY 2341 AATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAATGTCAAGTCTCAGCTTATAG 2400
Db 2340 AATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAATGTCAAGTCTCAGCTTATAG 2399
QY 2401 GACTGCCATGAAACTCCGAAGACTGCAGAAAGCCCTTTGCTTGGATCTCTTGAGCCTGTC 2460
Db 2400 GACTGCCATGAAACTCCGAAGACTGCAGAAAGCCCTTTGCTTGGATCTCTTGAGCCTGTC 2459
QY 2461 AGCTGCATGTGATGCCTTGGACCAACCACTCAAGCAAAATGACCAAGCCATGGATAT 2520
Db 2460 AGCTGCATGTGATGCCTTGGACCAACCACTCAAGCAAAATGACCAAGCCATGGATAT 2519
QY 2521 CCTGCAGATTATTAATGTGTGACCACTATTATGACCGCCTGGAGCAAGAGCACAACAA 2580
Db 2520 CCTGCAGATTATTAATGTGTGACCACTATTATGACCGCCTGGAGCAAGAGCACAACAA 2579
QY 2581 TTTGGTCAACGTCCTCTCTGCGTGGATATGTGTCTGAACCTGGCTGCTGAATGTTTATGA 2640
Db 2580 TTTGGTCAACGTCCTCTCTGCGTGGATATGTGTCTGAACCTGGCTGCTGAATGTTTATGA 2639
QY 2641 TACGGGACGAAACAGGAGGATCCGTCTCTTTTAAACCTGGCATCATTTCCCTGTG 2700
Db 2640 TACGGGACGAAACAGGAGGATCCGTCTCTTTTAAACCTGGCATCATTTCCCTGTG 2699
QY 2701 TAAAGCACATTTGGAAGACAAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGG 2760
Db 2700 TAAAGCACATTTGGAAGACAAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGG 2759
QY 2761 A 2761
Db 2760 A 2760

Search completed: April 25, 2003, 08:42:52
Job time : 528.159 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 07:58:12 ; Search time 2604.24 Seconds
(without alignments)
17170.339 Million cell updates/sec

Title: US-09-845-416-9_COPY_540_3300
Perfect score: 2761
Sequence: 1 gccagacatttactgga.....aagtggcaagttcaacagga 2761

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308016

Minimum DB seq length: 0
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_hic:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: gb_gss:*
- 18: em_gss_hum:*
- 19: em_gss_inv:*
- 20: em_gss_pln:*
- 21: em_gss_vrt:*
- 22: em_gss_fun:*
- 23: em_gss_mam:*
- 24: em_gss_mus:*
- 25: em_gss_other:*
- 26: em_gss_pro:*
- 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	874.6	31.7	1490	11 BC009242	
2	605.4	21.9	834	13 BI729851	BC009242 Homo sapi
3	495.2	17.9	750	13 BI730168	BI729851 603349511
4	486.6	17.6	579	9 AL121550	BI730168 603349711
5	457	16.6	757	9 AL556247	AL121550 DKF2p762L
6	446	16.2	620	14 BQ640063	AL556247 AL556247 BQ640063 he23g04.y

7	435.4	15.8	770	12	BG719710
8	424.2	15.4	663	13	BM488464
9	412.8	15.0	652	10	BB629984
10	412.8	15.0	2135	11	AK013510
11	411.2	14.9	704	10	BB610411
12	409.4	14.8	599	10	BB666688
13	395.4	14.3	854	9	AI196693
14	381.8	13.8	423	9	AA460476
15	346.6	12.6	684	9	AL641565
16	345.8	12.5	898	12	BF182065
17	345	12.5	633	9	AL796733
18	330	12.0	608	9	AL672616
19	329	11.9	532	10	BE334408
20	325.6	11.8	532	13	BJ036615
21	315.4	11.4	541	13	BI546771
22	310	11.2	488	13	BM488542
23	297.8	10.8	597	9	AL795652
24	293.8	10.6	696	13	BJ075057
25	276.6	10.0	647	9	AL652775
26	271.6	9.8	2334	11	BC011062
27	265.8	9.6	784	12	BG212445
28	265.6	9.6	665	10	BE201973
29	254	9.2	772	13	BI250598
30	250	9.1	502	9	AL602076
31	246	8.9	835	13	BI553820
32	230.4	8.3	573	13	BI446561
33	226	8.2	250	12	BF963618
34	206.4	7.5	479	14	BQ304046
35	199.2	7.2	733	12	BF144271
36	192	7.0	501	9	AA771140
37	184.8	6.7	307	12	BF226333
38	182.4	6.6	645	10	AA467977
39	178.2	6.5	600	13	BI988528
40	173	6.3	726	10	BB627285
41	172	6.2	646	13	BI289102
42	156.8	5.7	475	12	BG207912
43	153.4	5.6	481	10	BE557463
44	153.2	5.5	885	11	AK020881
45	151.8	5.5	417	12	BF387048

ALIGNMENTS

RESULT 1	BC009242	Homo sapiens, dystrophin (muscular dystrophy, Duchenne and Becker types), includes DXS142, DXS164, DXS206, DXS230, DXS239, DXS268, DXS269, DXS270, DXS272, clone IMAGE:3029414, mRNA.	1490 bp	mRNA	linear	HTC 12-JUL-2001
LOCUS	BC009242	HTC				
DEFINITION	BC009242	HTC				
ACCESSION	BC009242	GI:14714379				
VERSION	BC009242.1					
KEYWORDS	HTC					
SOURCE	Homo sapiens					
ORGANISM	Homo sapiens					
REFERENCE	1	(bases 1 to 1490)				
AUTHORS	Strausberg,R.					
TITLE	Direct Submission					
JOURNAL	Submitted (06-JUN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA					
REMARK	NIH-MGC Project URL: http://mgc.nci.nih.gov					
COMMENT	Contact: MGC help desk Email: cgapbs-r@mail.nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Ruben Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Institute for Systems Biology contact: amadan@systemsbiology.org					

Mon Apr 28 09:28:41 2003

Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia Greene, Mark Kettelman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov> Series: IRAL Plate: 10 Row: j Column: 10 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 5032280 This clone has the following problem: incomplete processing.

Location/Qualifiers

1. 1490

/organism="Homo sapiens"
/db_xref="LocustID:1756"
/db_xref="taxon:9606"
/clone="IMAGE:3029414"
/tissue_type="Muscle, rhabdomyosarcoma"
/clone_lib="NIH_MGC_17"
/lab_host="DH10B-R"
/note="Vector: pOTB7" 358 t

BASE COUNT 505 a 299 c 328 g 358 t

ORIGIN

Query Match 31.7%; Score 874.6; DB 11; Length 1490;
Best Local Similarity 99.5%; Pred. No. 3.3e-203;
Matches 877; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 GCCAGACCTATTGACTGGAATAGTGGTTTGGCAGCAGTCAGCCACACAAACGACTGGA 60
Db 609 GCCAGACCTATTGACTGGAATAGTGGTTTGGCAGCAGTCAGCCACACAAACGACTGGA 668
61 ACATGCATTCAACATCGCCAGATATCAATTAGGCATAGAGAACTACTCGATCCTGAAGA 120
Db 669 ACATGCATTCAACATCGCCAGATATCAATTAGGCATAGAGAACTACTCGATCCTGAAGA 728
121 TGTGATACCACTATCCAGATAAGAGTCCATCTTAATGTACATCATCATCTTCCA 180
QY 121 TGTGATACCACTATCCAGATAAGAGTCCATCTTAATGTACATCATCATCTTCCA 180
Db 729 TGTGATACCACTATCCAGATAAGAGTCCATCTTAATGTACATCATCATCTTCCA 788
181 AGTTTGGCTCAACAAGTGAGCATTTGAAGCCATCCAGGAAGTGGAAATGTTGCCAAGGCC 240
QY 181 AGTTTGGCTCAACAAGTGAGCATTTGAAGCCATCCAGGAAGTGGAAATGTTGCCAAGGCC 848
Db 789 AGTTTGGCTCAACAAGTGAGCATTTGAAGCCATCCAGGAAGTGGAAATGTTGCCAAGGCC 300
241 ACCTAAAGTGACTAAAGAAGAACATTTTCAGTTACATCATCAATGTCACATTTCTCAACA 300
QY 241 ACCTAAAGTGACTAAAGAAGAACATTTTCAGTTACATCATCAATGTCACATTTCTCAACA 908
Db 849 ACCTAAAGTGACTAAAGAAGAACATTTTCAGTTACATCATCAATGTCACATTTCTCAACA 360
301 GATCAGGTCAGTCTAGCACAGGGATATGAGAGAACTTCTCCCTTAAGCCTCGATTCAA 360
QY 301 GATCAGGTCAGTCTAGCACAGGGATATGAGAGAACTTCTCCCTTAAGCCTCGATTCAA 968
Db 909 GATCAGGTCAGTCTAGCACAGGGATATGAGAGAACTTCTCCCTTAAGCCTCGATTCAA 420
361 GAGCTATGCCTACACACAGGCTGCTTATGTACACACCTCTGACCCCTACACGAGCCCAT 420
QY 361 GAGCTATGCCTACACACAGGCTGCTTATGTACACACCTCTGACCCCTACACGAGCCCAT 1028
Db 969 GAGCTATGCCTACACACAGGCTGCTTATGTACACACCTCTGACCCCTACACGAGCCCAT 480
421 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAAGTTCATTGGCAGTTTCATTGGAGAG 480
QY 421 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAAGTTCATTGGCAGTTTCATTGGAGAG 1088
Db 1029 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAAGTTCATTGGCAGTTTCATTGGAGAG 540
481 TGAAGTAAACCTGGACCGTTATCAACAGCTTTAGAGAAGTATATCGTGGCTTCTTTC 540
QY 481 TGAAGTAAACCTGGACCGTTATCAACAGCTTTAGAGAAGTATATCGTGGCTTCTTTC 1148
Db 1089 TGAAGTAAACCTGGACCGTTATCAACAGCTTTAGAGAAGTATATCGTGGCTTCTTTC 600
541 TGCTGAGGACACATTCGAAGCACAAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 600
QY 541 TGCTGAGGACACATTCGAAGCACAAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 1208
Db 1149 TGCTGAGGACACATTCGAAGCACAAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 660
601 CCAGTTTCATCTCATGAGGGGTACATGATGGATTTGACAGCCCATCAGGCGGGTGG 660
QY 601 CCAGTTTCATCTCATGAGGGGTACATGATGGATTTGACAGCCCATCAGGCGGGTGG 1268
Db 1209 CCAGTTTCATCTCATGAGGGGTACATGATGGATTTGACAGCCCATCAGGCGGGTGG 720
661 TAATATTTACAAATGGGAAGTAAGCTGATTTGGAACAGGAAATTTATCAGAAGATGAAGA 720
QY 661 TAATATTTACAAATGGGAAGTAAGCTGATTTGGAACAGGAAATTTATCAGAAGATGAAGA 1328
Db 1269 TAATATTTACAAATGGGAAGTAAGCTGATTTGGAACAGGAAATTTATCAGAAGATGAAGA

QY 721 AACTGAAGTACAAAGAGCAGATGAATCTCTAAATTCAGATGGGAATGCCTCAGGGTAGC 780
Db 1329 AACTGAAGTACAAAGAGCAGATGAATCTCTAAATTCAGATGGGAATGCCTCAGGGTAGC 1388
QY 781 TAGCATGGAAAAACAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT 840
Db 1389 TAGCATGGAAAAACAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT 1448
QY 841 GAAAGAGTTGAATGACTGGCTAAACAAAACAGAAAGAAAGAA 881
Db 1449 GAAAGAGTTGAATGACTGGCTAAACAAAACAGAAAGAAAGAA 1489

RESULT 2
BI729851 834 bp mRNA linear EST 20-SEP-2001
LOCUS 603349511F1 NIH_MGC_94 Mus musculus cDNA clone IMAGE:5357162 5',
DEFINITION mRNA sequence.

ACCESSION BI729851
VERSION BI729851.1 GI:15706864

KEYWORDS EST.
SOURCE house mouse.
ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 834)
NIH-MGC <http://mgc.nci.nih.gov/>.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

Tissue Procurement: The Cepko Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
cDNA Sequencing by: Incyte Genomics, Inc.
DNA distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM1908 row: e column: 03
High quality sequence stop: 796.
Location/Qualifiers
1. 834
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5357162"
/clone_lib="NIH_MGC_94"
/tissue_type="retina"
/lab_host="DH10B (phage-resistant)"
/note="Organ: eye; Vector: pCMV-SPORT6; Site: 1; NotI;
Site: 2; SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 3.3 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH_MGC Library."

FEATURES
source

BASE COUNT 275 a 173 c 199 g 187 t
ORIGIN

Query Match 21.9%; Score 605.4; DB 13; Length 834;
Best Local Similarity 88.6%; Pred. No. 2.4e-137;
Matches 690; Conservative 0; Mismatches 86; Indels 3; Gaps 3;

QY 1088 TGGGCAACATCTGTAGATGGACAGAACCCGCTGGTCTTTTACAGACACTCATAGA 1147
Db 31 TGGGCAACATCTGTAGATGGACAGAACCCGCTGGTCTTTTACAGACACTCATAGA 90
QY 1148 TTAGTCAACAGTTCCCTCCCTGGACCTGGAAAAGTTTCTTGCCTGGCTTACAGAACTGAA 1207
Db 91 TTAGTCAACAGTTCCCTCCCTGGACCTGGAAAAGTTTCTTGCCTGGCTTACAGAACTGAA 150
QY 1208 ACAACTGCCAATGTCTTACAGGATGCTACCCGTAAGGAAAGGCTCTTAGAAGACTCCAG 1267
Db 151 ACAACTGCCAATGTCTTACAGGATGCTACCCGTAAGGAAAGGCTCTTAGAAGACTCCAG 209

QY	1268	GGAGTAAAGAGCTGATGAAACAATGGCAAGACCTCCAAGGTGAAATTGAAGCTCACACA	1327
Db	210	GGAGTCAGAGAGCTGATGAAACCATGGCAAGATCTCCAAGGAGAAATTCGAAACTCACACA	269
QY	1328	GATGTTTATCACAAACCTGGATGAAACAGCCAAAAAATCCTGAGATCCCTGGAAGGTTCC	1387
Db	270	GATATCTATCACAAATCTGATGAAAATGGCCAAAAAATCCTGAGATCCCTGGAAGGTTCC	329
QY	1388	GATGATGCAGTCCTGTTACAAAGACGTTTGGATAACATGAACCTCAAGTGGAGTGAACCT	1447
Db	330	GATGAAGCACCCCTGTTACAAAGACGTTTGGATAACATGAATTTCAAGTGGAGTGAACCT	389
QY	1448	CGGAAAAAGTCTTCAACATTAGGTCCCATTTGGAAGCCAGTTCTGACCAGTGGAAAGCGT	1507
Db	390	CAGAAAAAGTCTCTCAACATTAGGTCCCATTTGGAAGCAAGTTCTGACCAGTGGAAAGCGT	449
QY	1508	CTGCACCTTCTCTGCAGGAACCTTCTGGTGTGGCTACAGCTGAAAGATGATGAATTAAGC	1567
Db	450	TTGCATCTTCTCTTCAGGAACCTTCTTGTGTGGCTACAGCTGAAAGATGATGAATTAAGC	1627
QY	1568	CGCAGGCACCTATTGGAGGGGACTTTCCAGCAGTTCAGAAGCAGAACGATGATACATAGG	1627
Db	510	CGTCAGGCACCCATCGGTGGTGATTTCCAGCAGTTCAGAAGCAGAATGATATACATAGG	569
QY	1628	GCCTTCAAGAGGGGAATTGAAAACCTAAGAACCTGTAATCATGAGTACTCTTGAGACTGTA	1687
Db	570	GCCTTCAAGAGGGGAATTGAAAACCTAAGAACCTGTAATCATGAGTACTCTTGAGACTGTA	629
QY	1688	CGAATATTTCTGACAGAGCAGCCCTTTGGAAGGACTAGAGAACTC-TACCAGGAGCCAG	1746
Db	630	AGAATATTTCTGACAGAGCAGCCCTTTGGAAGGACTAGAGAACTC-TACCAGGAGCCAG	689
QY	1747	AGAGCTGCCCTCCTGAGGAGAGAGCCAGAAATGTCACTCGGCTCTTACGAAAGCAGGCTGA	1806
Db	690	AGAAGTGCCTCCTGAAAGAAAGAGCTCAGAAATGTCACTCGGCTCCTACGAAAGCAGGCTGA	749
QY	1807	GGAGGTCAATACTAGTGGGAAAAAATTGAACCTGCACCTC-CGCTGACTGGCAGAGAAAA	1864
Db	750	AGAGGTCAACGCTGAATGGGACCAATTGAACCTCGGCTCAAGCTGATTGGCAGAGAAAA	808

FEATURES
source

```

1. 1750
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:5357187"

```

RESULT 4
AL121550
LOCUS
DEFINITION

Accession	Length	Source	Library	EST	Accession
AL121550	579 bp	mRNA	linear	EST 25-FEB-2000	
DFE2p762L078_r1	762 (synonym: hm12)	Homo sapiens		cDNA clone	
DFE2p762L078_5'	mRNA sequence.				

```
/clone_lib="NIH_MGC_94"  
/tissue_type="retina"
```

```
/lab_host="DH10B (phage-resistant)"
```

```
/note="Organ: eye; Vector: pCMV-SPORT6; Site_1: NotI;  
Site 2: SalI; closed ..."
```

BASE COUNT	240 a	161 c	178 g	170 t	1 others
ORIGIN	Note: this is a NIH_MGC Library.				

Query Match 17.9%; Score 495.2; DB 13; Length 750;
Best Local Similarity 85.5%; Pred. No. 2.3e-110;
Matches 636; Conservative 0. Mismatched 0

RESULT 4
AL121550
LOCUS
DEFINITION

Mon Apr 28 09:28:41 2003

```

AL121550
AL121550.1 GI:5927551
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 579)
Ottewaelder, B., Obermaier, B., Mewes, H.W., Gassenhuber, J. and
Wiemann, S.
EST (Ottewaelder, et al.)
Unpublished (1999)
JOURNAL
COMMENT
Contact: Ottenwaelder B
MIPS
Am Klopferspitz 18a D-82152 Martinsried, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by MediGenomix (Martinsried/Germany) within the cDNA
sequencing consortium of the German Genome Project. No s1 sequence
available.
This clone (DKFZp762L078) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
Location/Qualifiers
1..579
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="DKFZp762L078"
/tissue_type="melanoma (MeWo cell line)"
/dev_stage="adult"
/lab_host="DH10B"
/note="Vector: pSPORT1; Site_1: NotI; Site_2: SalI"
BASE COUNT 180 a 128 c 144 g 127 t
ORIGIN
Query Match 17.6%; Score 486.6; DB 9; Length 579;
Best Local Similarity 97.2%; Pred. No. 2.7e-108;
Matches 495; Conservative 0; Mismatches 14; Indels 0; Gaps 0;
QY 1296 AAGACCTCCAAAGGTGAAATTGAAGCTCACACAGATGTTTATCACACCTGGATGAAACA 1355
DB 71 AGGACCTCCAAAGGTGAAATTGAAGCTCACACAGATGTTTATCACACCTGGATGAAACA 130
QY 1356 GCCAAAAAATCCTGAGATCCCTGGAAAGTTCCGATGATGCAGTCCCTGTTACAAAGACGTT 1415
DB 131 GCCAAAAAATCCTGAGATCCCTGGAAAGTTCCGATGATGCAGTCCCTGTTACAAAGACGTT 190
QY 1416 TGGATAACATGAACCTCAAGTGGAGTGAACCTCGGAAAAAGTCTCTCAACATTAGGTCCC 1475
DB 191 TGGATAACATGAACCTCAAGTGGAGTGAACCTCGGAAAAAGTCTCTCAACATTAGGTCCC 250
QY 1476 ATTTGGAAGCCAGTTCTGACCAAGTGGAAAGCGTCTGCACTTTCTCTGCAGGAACCTCTGG 1535
DB 251 ATTTGGAAGCCAGTTCTGACCAAGTGGAAAGCGTCTGCACTTTCTCTGCAGGAACCTCTGG 310
QY 1536 TGTGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGACCTATTGGAGCCGACTTTC 1595
DB 311 TGTGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGACCTATTGGAGCCGACTTTC 370
QY 1596 CAGCAGTTTCAAGAGCAGAACGATGTACATAGGCGCTTCAAGAGGGGAATTGAAACTAAAG 1655
DB 371 CAGCAGTTTCAAGAGCAGAACGATGTACATAGGCGCTTCAAGAGGGGAATTGAAACTAAAG 430
QY 1656 AACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGG 1715
DB 431 AACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGG 490
QY 1716 AAGGACTAGAGAAACTCTACCAGGAGCCAGAGAGCTGCCTCCTGAGAGAGAGCCCGA 1775
DB 491 AAGGCTAGAGAAACTCTACCAGGAGCCAGAGAGCTGCCTCCTGAGAGAGAGCCCGA 550

```

```

QY 1776 ATGTCACTCGGCTTCTACGAAAGCAGGCT 1804
DB 551 AATGTCACTCGCTTCTACGAAAGCAGGCT 579
AL556247
AL556247 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DK001YB17 5
prime, mRNA sequence.
AL556247
AL556247.1 GI:12898746
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 757)
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
JOURNAL
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1..757
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DK001YB17"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 190 a 197 c 188 g 182 t
ORIGIN
Query Match 16.6%; Score 457; DB 9; Length 757;
Best Local Similarity 100.0%; Pred. No. 5.3e-101;
Matches 457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2305 CCACGAGACTCAAAACAACCTGCTGGGACCATCCCAAAATGACAGAGCTCTACCACTCTT 2364
DB 128 CCACGAGACTCAAAACAACCTGCTGGGACCATCCCAAAATGACAGAGCTCTACCACTCTT 187
QY 2365 AGCTGACCTGAATAATGTGATCTCTGAGCCTGTGAGCCTGTGATGATGATGATGATGAT 2424
DB 188 AGCTGACCTGAATAATGTGATCTCTGAGCCTGTGAGCCTGTGATGATGATGATGATGAT 247
QY 2425 GCAGAAAGCCCTTTGCTTGGATCTCTGAGCCTGTGAGCCTGTGATGATGATGATGATGAT 2484
DB 248 GCAGAAAGCCCTTTGCTTGGATCTCTGAGCCTGTGAGCCTGTGATGATGATGATGATGAT 307
QY 2485 GCACAACCTCAAGCAAAATGACCAAGCAAGCAACAAATTTGGTCAACGCTCCTCTCTGCGT 2544
DB 308 GCACAACCTCAAGCAAAATGACCAAGCAAGCAACAAATTTGGTCAACGCTCCTCTCTGCGT 367
QY 2545 CACTATTTATGACCGCCTTGAGCAAGCAAGCAACAAATTTGGTCAACGCTCCTCTCTGCGT 2604
DB 368 CACTATTTATGACCGCCTTGAGCAAGCAAGCAACAAATTTGGTCAACGCTCCTCTCTGCGT 427
QY 2605 GGATATGTGTCTCAACTGGCTGCTGAATGTTTATGATACGGGACGAGGAGGATCCG 2664
DB 428 GGATATGTGTCTCAACTGGCTGCTGAATGTTTATGATACGGGACGAGGAGGATCCG 487

```



```
QY 2665 TGTCCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTA 2724
|||||
Db 488 TGTCCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTA 547
|||||
QY 2725 CAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGA 2761
|||||
Db 548 CAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGA 584
|||||

RESULT 6
BQ640063
LOCUS
DEFINITION
he23g04.y1 Human Retina cDNA (Un-normalized, unamplified): hd/he
Homo sapiens cDNA clone he23g04 5', mRNA sequence.
ACCESSION
BQ640063
VERSION
BQ640063.1 GI:21764522
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 620)
AUTHORS
Wistow,G., Bernstein,S.L., Wyatt,M.K., Ray,S., Behal,A., Touchman
,J.W., Bouffard,G., Smith,D. and Peterson,K.
TITLE
Expressed sequence tag analysis of human retina for the NEIBank
Project: Retbindin, an abundant, novel retinal cDNA and alternative
splicing of other retina-preferred gene transcripts
Mol. Vis. 8 (4), (2002) In press
JOURNAL
Contact: Wistow G
COMMENT
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 23 row: g column: 04
Seq primer: M13RP1 reverse primer (ABI).
Location/Qualifiers
1..620
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="he23g04"
/clone_lib="Human Retina cDNA (Un-normalized, unamplified
): hd/he"
/tissue_type="Retina"
/dev_stage="Adult"
/lab_host="EMDH10B"
/note="Organ: Eye; Vector: pSPORT1; Neural retina tissue
was dissected from two 80 year old donors with no observed
eye disease. 100ug of total RNA was used for library
construction. A directionally cloned cDNA library in the
pSPORT1 vector (Life Technologies) was constructed at
Bioserve Biotechnology (Laurel MD) essentially following
the protocols of the SuperScript Plasmid System full
details of which are contained in the manufacturer's
Instruction manual (http://www.lifetech.com/). First
strand synthesis was carried out using a Not I
primer-adaptor [5'-pGACTAGTCTAGATCGCGAGCGCGCCGCC(T)15-3'
]. EST analysis was performed on the unamplified library
at the NIH Intramural Sequencing Center (NISC)."
BASE COUNT 165 a 148 c 144 g 163 t
ORIGIN
Query Match
Best Local Similarity 16.2%; Score 446; DB 14; Length 620;
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2316 AAACAACCTGTCTGGGACCATCCCAAATGACAGAGCTCTACCAGTCTTAGCTGACCTGA 2375
|||||
Db 1 AAACAACCTGTCTGGGACCATCCCAAATGACAGAGCTCTACCAGTCTTAGCTGACCTGA 60
|||||
QY 2376 ATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAACTCCGAAGACTGCAGAAGGCC 2435
|||||
```

```
Db 61 ATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAACTCCGAAGACTGCAGAAGGCC 120
QY 2436 TTTGCTTGGATCTCTTGAGCCTGTGAGCTGCATGTGATGCTGCGTGGACCAACCTCA 2495
|||||
Db 121 TTTGCTTGGATCTCTTGAGCCTGTGAGCTGCATGTGATGCTGCGTGGACCAACCTCA 180
QY 2496 AGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTGACCACATATTATG 2555
|||||
Db 181 AGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTGACCACATATTATG 240
QY 2556 ACCGCCTGGAGCAAGAGCAACAATTTGGTCAACGCTCCCTCTCTGCGTGGATATGTGTC 2615
|||||
Db 241 ACCGCCTGGAGCAAGAGCAACAATTTGGTCAACGCTCCCTCTCTGCGTGGATATGTGTC 300
QY 2616 TGAACCTGGCTGCTGAATGTTTATGATACGGGACGAGGAGGATCCGTCCTGCTCTT 2675
|||||
Db 301 TGAACCTGGCTGCTGAATGTTTATGATACGGGACGAGGAGGATCCGTCCTGCTCTT 360
QY 2676 TTAACCACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAGATACCTTT 2735
|||||
Db 361 TTAACCACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAGATACCTTT 420
QY 2736 TCAAGCAAGTGGCAAGTTTCAACAGGA 2761
|||||
Db 421 TCAAGCAAGTGGCAAGTTTCAACAGGA 446

RESULT 7
BQ719710
LOCUS
DEFINITION
BQ719710 NIH_MGC_97 Homo sapiens cDNA clone linear EST 08-MAY-2001
mRNA sequence.
ACCESSION
BQ719710
VERSION
BQ719710.1 GI:13998897
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 770)
AUTHORS
NIH-MGC http://mgc.nci.nih.gov/.
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10731 row: 1 column: 08
High quality sequence stop: 767.
Location/Qualifiers
1..770
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4822807"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTTN-3',
size-selected for average insert size 2.2 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NIMH/NHGRI, National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 189 a 199 c 198 g 184 t
ORIGIN
```


/clone_lib="Normalized Chicken Breast Muscle, Leg Muscle, and Epiphyseal Growth Plate cDNA library (pgm2n)"
/sex="Male and Female"
/tissue_type="Breast muscle, leg muscle and epiphyseal growth plate"
/dev_stage="Breast, leg: Embryo(d19); post-hatch(1d,1,3,5,7,9,11 weeks); growth plate(1d,7d,14d post-hatch)"
/lab_host="E. coli EMDH10B"
/note="Vector: pCMVSPORT6; Library made from equivalent pools of total RNA isolated from each tissue (embryonic muscle 33.3%; juvenile muscle 33.3%; and epiphyseal growth plate 33.3% of the final RNA pool). Single pass sequencing from 5'-end"
209 a 132 c 172 g 144 t 6 others

BASE COUNT 209 a 132 c 172 g 144 t 6 others
ORIGIN
Query Match 15.4%; Score 424.2; DB 13; Length 663;
Best Local Similarity 78.1%; Pred. No. 5.6e-93;
Matches 518; Conservative 0; Mismatches 144; Indels 1; Gaps 1;

QY 1279 GCTGATGAAACAATGGCAAGACCTCCCAAGGTGAAATTTGAAGCTCACACAGATGTTTATCA 1338
Db 1 GCTCATGAAGCAGTGGCAGGATCTACAGGCAGAAATTTGATGCACATACATGACATCT-TNN 59
QY 1339 CAACCTGGATGAAACAGCAACAAATCCTGAGATCCCTGGAAGGTTCCGATGATGCAGT 1398
Db 60 CAACCTGGATGAAACAGCAACAAATCCTGAGATCCCTGGAAGGCTCAGAGGATGCTGT 119
QY 1399 CCTGTTACAAAGACGTTTGGATAACATGAATTTCAAGTGGAGTGAATTCGGAAGAAATC 1458
Db 120 CCTGTTGACAGAGACGTTTGGATAACATGAATTTCAAGTGGAGTGAATTTAGGAAGAAATC 179
QY 1459 TCTCAACATTAGTCCCATTTTGAAGCCAGTTCTGACCAGTGGAAAGCCTGTCACCTTTC 1518
Db 180 TCTAAACATTAGATCTCATTTTGAAGCCAGCAGACAGCCAGTGGAAAGCCTTACATCTCTC 239
QY 1519 TCTGCAGGAACCTTCTGGTGGCTACAGCTGAAAGATGATGAATTAAGCCGAGGCAACC 1578
Db 240 TCTTCAGGAACCTTTTGGCATGGCTGCAATTTGAAGGAGGATGAATTAAGCAAGCAAGCACC 299
QY 1579 TATTGGAGGCGACTTTCCAGCAGTTCAGAGAGCAACGATGTACATAGGCGCTTCAAGAG 1638
Db 300 CATTGGTGGAGATATTTCCACTGTGCGAAGAGCAGAAATGATGTTTCATAGGACTTTCAGAG 359
QY 1639 GGAATTTGAAACTAAAGAACCTGTAAATCATGAGTACTCTTGAGACTGTACGAATATTTCT 1698
Db 360 GGAGCTGAAACAAACAAACAACTGTTATCATGAATGCACCTGAGACTGTGCGACTCTTCTC 419
QY 1699 GACAGAGCAGCCTTTTGAAGAGGACTAGAGAAACTCTACCAGAGCCAGAGAGTGCCTCC 1758
Db 420 GGCAGATCAACCCAGTAGAGGAGTGGAAAGGTTCTATCCAGAACCAAGAGACCTATCACC 479
QY 1759 TGAGGAGAGAGCCAGAAATGTCACCTGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATAC 1818
Db 480 TGAGGAGAGAGGCGCCAGAAATGTCACCTGGCTGACTGGCAGAGAAATATAGATGAGACCT 1878
QY 1819 TGAGTGGGAAAAATTTGAACCTGCACTCCGCTGACTGGCAGAGAAATATAGATGAGACCT 1938
Db 540 TGAGTGGGATAAGCTAAATCTACGTTCTGCTGATGGCAAGAAAGATAGATGATGCTCT 599
QY 1879 TGAAGACTCCAGGAACCTTCAAGAGGCGCAGGATGAGCTGGACCTCAAGCTGCGCCCAAGC 1938
Db 600 TGAAGACTGCAGGGTCTTTCAGGAGGCAATGGATGAACCTNNNCCTGAAACTGCGCCAGGC 659
QY 1939 TGA 1941
Db 660 TGA 662

RESULT 9
BB629984
LOCUS
DEFINITION BB629984 RIKEN full-length enriched, adult female vagina Mus
652 bp mRNA linear EST 31-AUG-2001

Query Match 15.8%; Score 435.4; DB 12; Length 770;
Best Local Similarity 98.3%; Pred. No. 1.1e-95;
Matches 451; Conservative 0; Mismatches 6; Indels 2; Gaps 1;

QY 2305 CCACGAGACTCAAAACAATGCTGGGACCATCCCAAAATGACAGAGCTTACCAGTCTTT 2364
Db 150 CCACGAGACTCAAAACAATGCTGGGACCATCCCAAAATGACAGAGCTTACCAGTCTTT 209
QY 2365 AGCTGACCTGAATAATGTGAGATCTCAGCTTATAGGACTGCCATGAACTCCGAAGACT 2424
Db 210 AGCTGACCTGAATAATGTGAGATCTCAGCTTATAGGACTGCCATGAACTCCGAAGACT 269
QY 2425 GCAGAAAGGCCCTTGTGATCTCTTGAGCTGTGATGATGATGATGATGATGATGATGAT 2484
Db 270 GCAGAAAGGCCCTTGTGATCTCTTGAGCTGTGATGATGATGATGATGATGATGATGAT 329
QY 2485 GCACAA--CCTCAAGCAAAATGACCCATGATATCCTGATATCCTGATATTTAATTTGTTG 2542
Db 330 GCACAAATCCTCAAGTCAGAAATGACCCATGATATCCTGATATCCTGATATTTAATTTGTTG 389
QY 2543 ACCACTATTATGACCGCTGGAGCAAGAGCACAACAATTTGGTCAACGTCCTCTCTGC 2602
Db 390 ACCACTATTATGACCGCTGGAGCAAGAGCACAACAATTTGGTCAACGTCCTCTCTGC 449
QY 2603 GTGGATATGTGCTGAACCTGGCTGCTGAATGTTATGATACGGGACGAACAGGGAGGATC 2662
Db 450 GTGGATATGTGCTGAACCTGGCTGCTGAATGTTATGATACGGGACGAACAGGGAGGATC 509
QY 2663 CGTGCTCTGCTTTTAAACACTGGCATCATTTCCCTGTGTAAAGACATTTGGAAGACAAG 2722
Db 510 CGTGCTCTGCTTTTAAACACTGGCATCATTTCCCTGTGTAAAGACATTTGGAAGACAAG 569
QY 2723 TACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGA 2761
Db 570 TACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGA 608

RESULT 8
BM488464
LOCUS
DEFINITION
pkm2n.pk007.117 Normalized Chicken Breast Muscle, Leg Muscle, and Epiphyseal Growth Plate cDNA library (pgm2n) Gallus gallus CDNA clone pgm2n.pk007.117 5', similar to sp|P11533|DMD_CHICK Dystrophin p1r|S02041 dystrophin, muscle - chicken emb|CAA31746.1| (X13369) dystrophin (AA 1 - 3660) [Gallus gallus], mRNA sequence.

ACCESSION BM488464
VERSION BM488464.1 GI:18609395
KEYWORDS EST.
SOURCE chicken.
ORGANISM Gallus gallus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae; Gallus.
1 (bases 1 to 663)
Cogburn, L.A. and Monsonego-Ornan, E.
ESTs from Normalized Chicken Breast Muscle, Leg Muscle, and Epiphyseal Growth Plate cDNA library, USDA/IFAFS Animal Genome Project

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
Unpublished (2002)
Contact: Larry A. Cogburn
University of Delaware
Townsend Hall, Newark, DE 19717, USA
Tel: 302-831-1335
Fax: 302-831-2822
Email: cogburn@udel.edu, www.chickest.udel.edu.
Location/Qualifiers
1. .663
/organism="Gallus gallus"
/strains="Commercial broiler and Ottawa Res. Centre Strains 90 & 21"
/db_xref="taxon:9031"
/clone="pgm2n.pk007.117"

FEATURES
source
/organism="Gallus gallus"
/strains="Commercial broiler and Ottawa Res. Centre Strains 90 & 21"
/db_xref="taxon:9031"
/clone="pgm2n.pk007.117"

BASE COUNT
202 a 232 c 208 g 209 t 3 others

/note="Organ: liver; Vector: pME18S-FL3; Site_1: DraIII (CACTGTGTG); Site_2: DraIII (CACCATGTG); 1st strand CDNA was primed with an oligo(dT) primer [ATGTGGCCTTTT TTTT TTTT TTTT]; double-stranded cDNA was ligated to a DraIII adaptor [TGTTGGCCTACTGG], digested and cloned into distinct DraIII sites of the pME18S-FL3 vector (5' site CACTGTGTG, 3' site CACCATGTG). XhoI should be used to isolate the cDNA insert. Size selection was performed to exclude fragments <1.5kb. Library constructed by Dr. Sumio Sugano (University of Tokyo Institute of Medical Science). Custom primers for sequencing: 5' end primer CTTTGTGCTCTAAAGCTGCG and 3' end primer CGACCTGCAGCTCGAGCACAC."

Query Match	14.3%;	Score 395.4;	DB 9;	Length 854;
Best Local Similarity	93.0%;	Pred. NO. 6.9e-86;		
Matches 425; Conservative	0;	Mismatches 31;	Indels 1;	Gaps 1;

QY	2305	CCACGAGACTCAAAACAACCTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAAGTCTTT	2364
Db	135	CCACGAGACCCAAACCACTTGTGGGACCAACCCCAAAATGACAGAGCTCTACCAAGTCTTT	194
QY	2365	AGCTGACCTGAATAATGTCCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	2424
Db	195	AGCTGACCTGAATAATGTCCAGTTCTCCGGTATAGGACTGCCATGAAAGCTCAGAAGGCT	254
QY	2425	GCAGAAGGCCCTTTTGCTTGGATCTCTTGAGCCCTGTCAGCTGCATGTGATGCCCTGGACCA	2484
Db	255	CCAGAAGGCCCTTTTGCTTGGATCTCTTGAGCCCTGTCAGCTGCATGTGATGCCCTGGACCA	314
QY	2485	GCACAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTGAC	2544
Db	315	GCACAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATAATTAACGTGTTGAC	374
QY	2545	CACATATTATGACCGCCTGGAGCAAGACGACACAACAAATTGGTCAACGTCCCTCTCTGCGT	2604
Db	375	TACAATTATGATCGTCTGGAGCAAGACGACACAACAAATCGGTCAATGTCCCTCTCTGTGT	434
QY	2605	GGATATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCC-	2663
Db	435	GGATATGTGTCTCAACTGGCTTCTCAATGTTTATGATACGGGACGAACAGGGAGGATCCG	494
QY	2664	GTGTCCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGT	2723
Db	495	GGGTCCTGCCCTTTTAAACTGGCATCATTTCTCTGTGTAAAGCACACTTGAAGACAAGT	554
QY	2724	ACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGG	2760
Db	555	ACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACTGG	591

RESULT 14	AA460476	LOCUS	AA460476	423 bp	mrna	linear	EST 09-JUN-1997
DEFINITION	zx6le10.r1 Soares_total_fetus.Nb2HF8_9w Homo sapiens cDNA clone IMAGE:795978 5', similar to gb:M18533 DYSTROPHIN (HUMAN);, mRNA sequence.	ACCESSION	AA460476	VERSION	AA460476.1	GI:2185222	
KEYWORDS	EST.	SOURCE	human.	ORGANISM	Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 423)	AUTHORS	Hillier,L., Allen,M., Bowles,L., Dubuque,T., Geisel,G., Jost,S., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.	TITLE	WashU-Merck EST Project 1997	Unpublished (1997)	

COMMENT

Contact: Willson RK
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: estewartson.wustl.edu

Email: est@watson.wustl.edu

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28ml3 rev2 ET from Amersham
High quality sequence stop: 287.

FEATURES

source

```

location/qualifiers
1. .423
/organism="Homo sapiens"
/db_xref="GDB:603969"
/db_xref="taxon:9606"
/clone="IMAGE:795978"
/clone_lib="Soares_total"
/dev_stage="8-9 weeks"
/lab_host="DH10B"

```

/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
 was prepared from mRNA obtained from pooled 8-9 week
 (total) fetus material with a Not I - oligo(dT) primer [5'
 TGTTACCAATCTGAAGTGGGAGCGCGCTTAATTTTTTTTTTTT 3'].
 Double-stranded cDNA was ligated to Eco RI adaptors
 (Pharmacia), digested with Not I and cloned into the Not I
 and Eco RI sites of the modified pT7T3 vector. Library
 went through one round of normalization, and was
 constructed by Bento Soares and M. Fatima Bonaldo. "
 107 c 98 g 104 t

BASE COUNT	114 a	107 c	98 g	104 t
ORIGIN	constructed by Bento soare			

Query Match 13.8%; Score 381.8; DB 9; Length 423;
Best Local Similarity 28.2

Matches	394;	Conservative	0;	Mismatches	2;	Indels	1;	Gaps	1;
QY	2305	CCACGAGACTCAAACAACCTTGCTGGGACCATTCCCAAAATGACAGAGCTCTACCCAGTCTTT	2364						
Db	27	CCACGAGACTCAAACAACCTTGCTGGGACCATTCCCAAAATGACAGAGCTCTACCCAGTCTTT	86						
QY	2365	AGCTGACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	2424						
Db	87	AGCTGACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	146						
QY	2425	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCTTGGACCA	2484						
Db	147	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCTTGGACCA	206						
QY	2485	GCACAACCTCAAGCAAAATGACCAAGCCCATGGATATCCTGCAGATTATTAAATTGTTTGAC	2544						
Db	207	GCACAACCTCAAGCAAAATGACCAAGCCCATGGATATCCTGCAGATTATTAAATTGTTTGAC	266						
QY	2545	CACATATTTATGACCGCTGGAGCAAGAGCACAACTTTGGTCAACGTCCTCTCTCGCT	2604						
Db	267	CACATATTTATGACCGCTGGAGCAAGAGCACAACTTTGGTCAACGTCCTCTCTCGCT	326						
QY	2605	GGATATGTCTGAACTGGCTGCTGAAT-GTTTATGATACGGGACGAACAGGGAGGATCC	2663						
Db	327	GGAGATGTCTGAACTGGCTGCTGAATAGTTTATGATACGGGACGAACAGGGAGGATCC	386						
QY	2664	GTGCTCTGCTTTTAAAACTGGCATCATTTCCCTGTG	2700						
Db	387	GTGCTCTGCTTTTAAAAACCGGCATCATTTCCCTGTG	423						

RESULT 15
AL641565

LOCUS AL641565

LOCUS	AL641565
DEFINITION	AL641565 YCC-8

DEFINITION AL641565 XGC-m
mRNA 500000000

ACCESSION	F06789
mRNA sequence.	

AL641565

VERSION AL641565.1 GI:16793690

KEYWORDS EST.

684 bp mRNA linear EST 12-DEC-2001
urana tropicalis cDNA clone Tneu012d20 5',

684 bp mRNA linear EST 12-DEC-2001
urana tropicalis cDNA clone Tneu012d20 5',

684 bp mRNA linear EST 12-DEC-2001
urana tropicalis cDNA clone Tneu012d20 5',

SOURCE western clawed frog.
ORGANISM Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae;
Xenopodinae; Silurana.
REFERENCE 1 (bases 1 to 684)
AUTHORS Huckle, E., Taylor, R., Ashurst, J.L., Zorn, A.M. and Rogers, J.
TITLE Sanger Xenopus tropicalis EST project 2001 (10_2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Huckle E
Sanger Centre
Hinxtton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
TROPICALIS_SEQUENCE_ID: TNeu012d20.sp6
Sequencing primer: SP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
FEATURES
 source
 1..684
 /organism="Silurana tropicalis"
 /db_xref="taxon:8364"
 /clone="TNeu012d20"
 /clone_lib="XGC-neurula"
 /dev_stage="neurula"
 /lab_host="Escherichia coli DH10B"
 /note="Vector: PCS107; Site_1: EcoRI; Site_2: NotI; cDNA
 was oligo dt primed from 5ug of poly A+ RNA from neurula.
 EcoRI-NotI cut cDNA was then ligated into PCS107 with
 EcoRI at the 5' end and NotI at the 3' end."
BASE COUNT 168 a 175 c 172 g 169 t
ORIGIN
Query Match 12.6%; Score 346.6; DB 9; Length 684;
Best Local Similarity 84.9%; Pred. No. 5.8e-74;
Matches 388; Conservative 0; Mismatches 69; Indels 0; Gaps 0;
QY 2305 CCACGAGACTCAAAACAACCTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAAGTCTTT 2364
 ||||| ||||| || ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 175 CCATGAGACACAAACACCTGCTGGGATCATCCCAAAATGACAGAAATTATACCAATCTTT 234
QY 2365 AGCTGACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT 2424
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 235 AGCTGACCTGAACAATGTCGATTTCTCAGCATACAGAACTGCCATGAAGCTAAGGAGATT 294
QY 2425 GCAGAAGGCCCTTTGCTGGATCTCTTGAGCCTGTCAGCTGCATGTCATGCCCTTGGACCA 2484
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 295 GCAAAAGGCCCTTGCTGGATTGCTAGGGCTGCTGCGAGCTTGGAAGCCTTGGACCA 354
QY 2485 GCACAACTCAAGCAAAATGACCAAGCCCATGGATATCCTGCAGATTATTAATTGTTGAC 2544
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 355 GCACAACTGAAGCAGATGACCAAGCTGATGGACATCCTGCAGATTATTAATTGCTTGAC 414
QY 2545 CACTATTATGACCGCTGGAGCAAGAGCACAACAATTGGTCAACGTCCTCTCTGCGT 2604
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 415 CACAATTATGATCGACTGGAGCAAGAGCACAATAATCTGGTGAACGTTCTCTCTGCGT 474
QY 2605 GGATATGTGTGAAGTGGCTGGTGAATGTTTATGATACGGGACGAACAGGAGGATCCG 2664
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 475 GGACATGTGCTCAACTGGCTGCTGAATGTTTATGACACGGGTGCAACGGGACGTATACG 534
QY 2665 TGTCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTA 2724
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 535 CGTCTATCTTTTAAACTGGTGTAAATTTCCCTGTGTAAAGCACATTTGGAAGATAAGTA 594
QY 2725 CAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGA 2761
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 595 CAGATACTTATCAAGCAGGTGGCAAGCTCCACGGGA 631

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 08:10:12 ; Search time 62.8837 Seconds
(without alignments)
13465.096 Million cell updates/sec

Title: US-09-845-416-9_COPY_540_3300
Perfect score: 2761
Sequence: 1 gccagacctatttgactgga.....aagtggaagttcaacagga 2761

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 875442

Minimum DB seq length: 0
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	40.4	1.5	2223	1	US-08-257-073-4
2	38.6	1.4	1995	1	US-08-425-069-3
3	38.6	1.4	1995	2	US-08-317-844B-3
4	38.2	1.4	1131	6	5180810-3
5	38.2	1.4	1784	6	5180810-2
6	38	1.4	1394	4	US-09-247-155-76
7	36.8	1.3	1886	6	5210183-1
8	36.6	1.3	1845	4	US-08-887-534A-22
9	36.2	1.3	289	4	US-09-007-005-17
10	36.2	1.3	289	4	US-09-244-796-17
11	36.2	1.3	2447	2	US-09-014-969-14
12	36	1.3	2873	4	US-08-630-915A-193
13	35.8	1.3	790	4	US-09-461-474-13
14	35.6	1.3	4868	1	US-08-139-937-12
15	35.6	1.3	4868	5	PCT-US93-11310-12
16	35.2	1.3	3466	1	US-08-468-036-38
17	35.2	1.3	3466	2	US-08-376-843-38
18	35	1.3	2169	4	US-09-434-408-3
19	35	1.3	4766	5	PCT-US93-07261-10
20	34.8	1.3	2442	4	US-09-040-485-1
21	34.8	1.3	3337	1	US-08-072-610-1
22	34.8	1.3	3337	2	US-08-719-822B-1
23	34.8	1.3	3337	4	US-09-092-458-1
24	34.6	1.3	198	5	PCT-US95-10668-3
25	34.6	1.3	198	5	PCT-US95-10668-4
26	34.6	1.3	2277	1	US-08-676-967-5
27	34.6	1.3	2277	1	US-08-676-974-5

c 28	34.6	1.3	2277	2	US-09-098-487-5	Sequence 5, Appli
c 29	34.6	1.3	2704	4	US-08-857-076-44	Sequence 44, Appli
c 30	34.6	1.3	3499	4	US-08-857-076-43	Sequence 43, Appli
31	34.4	1.2	717	4	US-09-461-697-189	Sequence 189, App
32	34.4	1.2	774	4	US-09-461-697-187	Sequence 187, App
33	34.4	1.2	819	4	US-09-461-697-185	Sequence 185, App
34	34.4	1.2	1047	4	US-09-671-950-1	Sequence 1, Appli
35	34.4	1.2	1047	4	US-09-671-950-3	Sequence 3, Appli
36	34.4	1.2	1047	4	US-09-671-950-5	Sequence 5, Appli
37	34.4	1.2	1047	4	US-09-671-950-7	Sequence 7, Appli
38	34.4	1.2	1047	4	US-09-671-950-9	Sequence 9, Appli
39	34.4	1.2	1047	4	US-09-671-950-11	Sequence 11, Appli
40	34.4	1.2	1047	4	US-09-671-950-13	Sequence 13, Appli
41	34.4	1.2	1669	4	US-09-461-697-184	Sequence 184, App
c 42	34.2	1.2	621	4	US-09-328-111-204	Sequence 204, App
c 43	34.2	1.2	734	3	US-09-014-583-1	Sequence 1, Appli
44	34.2	1.2	750	4	US-08-961-527-370	Sequence 370, App
45	34.2	1.2	1089	3	US-08-642-807A-27	Sequence 27, Appli

ALIGNMENTS

RESULT 1
US-08-257-073-4
; Sequence 4, Application US/08257073
; Patent No. 5766597
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: de Taisne, Charles
; APPLICANT: Tine, John A.
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: New York
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,073
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/075,783
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/852,305
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,183
; FILING DATE: 20-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2570
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066 CURTMS
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2223 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-257-073-4

Mon Apr 28 09:28:39 2003

Query Match 1.5%; Score 40.4; DB 1; Length 2223;
Best Local Similarity 51.7%; Pred. No. 0.056; Indels 0; Gaps 0;
Matches 92; Conservative 0; Mismatches 86;

828 AGAATCAGAACTGAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGAAACAAGGA 887
2033 AAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2092
888 AAATGGAGGAAGAGCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCAAGTACAACAAC 947
2093 AAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 2152
948 ATAAAGGTGCTTCAAGAAGATCTAGAACAAGAACAAAGTCAGGGTCAATTTCTCACTCA 1005
2153 AAGAAGAAGAAACAAGAAAGAAAGAAAGAAAGAAAGAAAGAAATTTGACAACTGA 2210

RESULT 2
US-08-425-069-3
; Sequence 3, Application US/08425069
; Patent No. 5728810
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Xu, Ming
; APPLICANT: Hinman, Michael B.
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 301 No. 5728810th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/425,069
; FILING DATE: 19-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-106P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1995 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; IMMEDIATE SOURCE:
; CLONE: p6B
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1785

US-08-425-069-3
; Sequence 3, Application US/08425069
; Patent No. 5728810
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Xu, Ming
; APPLICANT: Hinman, Michael B.
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 301 No. 5728810th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/425,069
; FILING DATE: 19-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-106P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1995 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO
; IMMEDIATE SOURCE:
; CLONE: p6B
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1785

Query Match 1.4%; Score 38.6; DB 1; Length 1995;
Best Local Similarity 44.0%; Pred. No. 0.19; Indels 0; Gaps 0;
Matches 164; Conservative 0; Mismatches 209;

US-08-425-069-3
; Sequence 3, Application US/08317844B
; Patent No. 5989894
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Xu, Ming
; APPLICANT: Hinman, Michael B.
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 301 No. 5989894th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,844B
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-105P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 241-1300
; TELEFAX: (703) 241-2848
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1995 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO

US-08-425-069-3
; Sequence 3, Application US/08317844B
; Patent No. 5989894
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Xu, Ming
; APPLICANT: Hinman, Michael B.
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 301 No. 5989894th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,844B
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-105P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 241-1300
; TELEFAX: (703) 241-2848
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1995 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO

US-08-425-069-3
; Sequence 3, Application US/08317844B
; Patent No. 5989894
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Xu, Ming
; APPLICANT: Hinman, Michael B.
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 301 No. 5989894th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,844B
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-105P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 241-1300
; TELEFAX: (703) 241-2848
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1995 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO

US-08-425-069-3
; Sequence 3, Application US/08317844B
; Patent No. 5989894
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Xu, Ming
; APPLICANT: Hinman, Michael B.
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 301 No. 5989894th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,844B
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-105P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 241-1300
; TELEFAX: (703) 241-2848
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1995 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO

US-08-425-069-3
; Sequence 3, Application US/08317844B
; Patent No. 5989894
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Xu, Ming
; APPLICANT: Hinman, Michael B.
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 301 No. 5989894th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,844B
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-105P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 241-1300
; TELEFAX: (703) 241-2848
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1995 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO

QY 1598 GCAGTTTCAGAAACGACGATGTACATAGGGCTTCAAGAGGGAATTTGAAAACTAAAGAA 1657
Db 731 GCAGTGCAGCTGCAGCAGCCGCGAGGACCTGGACAAACAGGACCCCGGAGGATATGGAC 790
QY 1658 CCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGAA 1717
Db 791 CAGGACAAACAGGACCATCTGGACCCGGTAGTCCCTGCGAGCAGCAGCCCGCGCAGCAG 850
QY 1718 GGACTAGAGAACTCTACAGAGGAGCCGAGAGAGCTGCTCTCTGAGGAGAGAGCCCAAGAT 1777
Db 851 GACCTGGAGGATATGGCCCTGGACAAACAGGACCCCGGAGGATATGGACCAACAAAG 910
QY 1778 GTCACCTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTGAAC 1837
Db 911 GACCATCTGGAGCAGGCTGAGGAGGTGAGGAGGTGAGGAGGTGAGGAGGTGAGGAGGTGAGG 970
QY 1838 CTGCACTCCGCTGACTGGCAGAGAAAAAATAGATGAGACCCCTTGAAGAGACTCCAGGAAC 1897
Db 971 TAGGAGGTTATGGACCAAGGACCAACAGGTCCTCAGGAGGATATGGACCAAGGACCAAGGTC 1030
QY 1898 CAAGAGCCACGGATGAGCTGGACCTCAAGCTCGCCCAAGCTGAGGTGATCAAGGGATCC 1957
Db 1031 CAGGAGGATATGGACCAAGGTCCTCAGGAGGATATGGACCAAGGACCAAGGTC 1090
QY 1958 TGGCAGCCCGGTGG 1970
Db 1091 AAGGACCAAGGAG 1103

RESULT 3
US-08-317-844B-3
; Sequence 3, Application US/08317844B
; Patent No. 5989894
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Xu, Ming
; APPLICANT: Hinman, Michael B.
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 301 No. 5989894th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,844B
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-105P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 241-1300
; TELEFAX: (703) 241-2848
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1995 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHETICAL: NO


```

; IMMEDIATE SOURCE:
; CLONE: p6B
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1785
US-08-317-844B-3

Query Match
Best Local Similarity 1.4%; Score 38.6; DB 2; Length 1995;
Matches 164; Conservative 0; Mismatches 209; Indels 0; Gaps 0;

QY 1598 GCAGTTTCAGAGCAGAACGATGTACATAGGCGCTTCAAGAGGGAATTTGAAAACTAAAGAA 1657
Db 731 GCAGTGCAGCTGCAGCAGCCCGCAGCAGGACCTGGACAACAAGGACCCGGAGGATATGGAC 790

QY 1658 CCTGTATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGAA 1717
Db 791 CAGGACAACAAGGACCATCTGGACCCCGTAGTGCCGCTGCAGCAGCAGCCCGCCGACGAG 850

QY 1718 GGACTAGAGAAACTCTACAGGAGCCCGCAGAGAGCTGCCTCCTGAGGAGAGAGCCCGAAT 1777
Db 851 GACCTGGAGGATATGGCCCTGGACAACAAGGACCCGGAGGATATGGACCAAGCAACAAG 910

QY 1778 GTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGAAAAATTGAAC 1837
Db 911 GACCATCTGGACAGGCGAGTGCAGCAGCAGCAGCCGCGCAGCAGGACCTGGACAACAAGGAT 970

QY 1838 CTGCACCTCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGAGACTCCAGGAATT 1897
Db 971 TAGGAGGTTATGGACCAAGCAACAAGGTCCAGGAGGATATGGACCAAGCAACAAGGTC 1030

QY 1898 CAAGAGGCCCGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGATCC 1957
Db 1031 CAGGAGGATATGGACCAAGCAACAAGGTCCAGGAGGATATGGACCAAGCAACAAGGTC 1090

QY 1958 TGGCAGCCCGTGG 1970
Db 1091 AAGGACCAAGGAGG 1103

RESULT 4
5180810-3
;Patent No. 5180810
; APPLICANT: Gomi, Hideyuki;Hozumi, Tatsunobu;Hattori, Shizuo;
; Tagawa, Chiaki;Kishimoto, Fumitaka;Bjorck, Lars
; TITLE OF INVENTION: PROTEIN H CAPABLE OF BINDING TO IGG
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/376.641
; FILING DATE: 07-JUL-1989
; SEQ ID NO:3
; LENGTH: 1131
5180810-3

Query Match
Best Local Similarity 1.4%; Score 38.2; DB 6; Length 1131;
Matches 130; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 701 AAATTATCAGAAGATGAAGAACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGA 760
Db 157 AAAAAGTTAGACGAAGATAATGCTAAACTTGTGAGGTTGTGAAACCAACAAGTTTGGAA 216

QY 761 TGGGAATGCCTCAGGTAAGTATGATGGAAGAAACAAAGCAATTTACATAGAGTTTAAATG 820
Db 217 AACGAAAACTCAAGAGTGAAGATGAGGAGATAAGAAAAATTTAGACAAACTTAGCAAA 276

QY 821 GATCTCCAGAAATCAGAAACTGAAAGAGTTGAATGACTGGCTTAACAAAAACAGAAGAA 880
Db 277 GAAATCAAGGAAAGCTCGAAAAATTTGGAGCTTGACTATCTCAAAAAATTTAGATCAGCAG 336

QY 881 ACAAGGAAATGGAGGAAGAGCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCAAGTA 940
Db 337 CACAAAGAGCACCACCAAAAGACCAACAAGAGAGAGCGGACAAAAAATCAAGAACA 396

```

```

QY 941 CAACAACATAAAGGTGCTTCAAGAAGATCTAGAACAAAGAACAG 983
Db 397 TTAGAACGTAAATACCAACGAGAGTAGAAAAACGTTTATCAAG 439

RESULT 5
5180810-2
;Patent No. 5180810
; APPLICANT: Gomi, Hideyuki;Hozumi, Tatsunobu;Hattori, Shizuo;
; Tagawa, Chiaki;Kishimoto, Fumitaka;Bjorck, Lars
; TITLE OF INVENTION: PROTEIN H CAPABLE OF BINDING TO IGG
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/376.641
; FILING DATE: 07-JUL-1989
; SEQ ID NO:2
; LENGTH: 1784
5180810-2

Query Match
Best Local Similarity 1.4%; Score 38.2; DB 6; Length 1784;
Matches 130; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 701 AAATTATCAGAAGATGAAGAACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGA 760
Db 484 AAAAAGTTAGACGAAGATAATGCTAAACTTGTGAGGTTGTGAAACCAACAAGTTTGGAA 543

QY 761 TGGGAATGCCTCAGGTAAGTACCATGGAAGAAACAAAGCAATTTACATAGAGTTTAAATG 820
Db 544 AACGAAAAACTCAAGAGTGAAGATGAGGAGATAAGAAAAATTTAGACAAACTTAGCAAA 603

QY 821 GATCTCCAGAAATCAGAAACTGAAAGAGTTGAATGACTGGCTTAACAAAAACAGAAGAA 880
Db 604 GAAATCAAGGAAAGCTCGAAAAATTTGGAGCTTGACTATCTCAAAAAATTTAGATCAGCAG 663

QY 881 ACAAGGAAATGGAGAGAGCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCAAGTA 940
Db 664 CACAAAGAGCACCACCAAAAGAACAAACAAGAACAAAGAGCGGACAAAAAATCAAGAACA 723

QY 941 CAACAACATAAAGGTGCTTCAAGAAGATCTAGAACAAAGAACAG 983
Db 724 TTAGAACGTAAATACCAACGAGAGTAGAAAAACGTTATCAAG 766

```

```

RESULT 6
US-09-247-155-76/c
; Sequence 76, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET.021A
; CURRENT APPLICATION NUMBER: US/09/247,155A
; CURRENT FILING DATE: 1999-02-09
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-08-10
; EARLIER APPLICATION NUMBER: 60/099,273
; EARLIER FILING DATE: 1998-10-04
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patent.pm
; SEQ ID NO 76
; LENGTH: 1394
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS

```


Mon Apr 28 09:28:39 2003

LOCATION: 46..675
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 46..87
OTHER INFORMATION: Von Heijne matrix
OTHER INFORMATION: score 5.9
OTHER INFORMATION: seq LTLGLSLILAGL/IV
FEATURE:
NAME/KEY: polyA_signal
LOCATION: 1363..1368
FEATURE:
NAME/KEY: polyA_site
LOCATION: 1382..1394
US-09-247-155-76

Query Match 1.4%; Score 38; DB 4; Length 1394;
Best Local Similarity 59.1%; Pred. No. 0.22;
Matches 65; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 789 AAAAAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGATCAGAAACTGAAAGAGT 848
Db 1087 AAAAAAATTTAAATATTTTCAATTTTAAATTTGCTGCACCATAGATGAATAAGAGC 1028
QY 849 TGAATGACTGGCTAAACAAAAACAGAAAGAAACAAAGGAAATGGAGGAA 898
Db 1027 TTACTTAAAGGAAAGAAAGAAACAAACAAACAAACAAACAAACAAACAAACGATGGA 978

RESULT 7
5210183-1
;Patent No. 5210183
; APPLICANT: LINDAHL, GUNNAR;FRITHZ, ELISABET;HEDEN, LARS-OLOF
; TITLE OF INVENTION: PROTEIN ARP, WITH IMMUNOGLOBULIN A
; BINDING ACTIVITY, THE CORRESPONDING VECTORS AND HOSTS, REAGENT
; KIT AND PHARMACEUTICAL COMPOSITION
; NUMBER OF SEQUENCES: 3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/270,099
; FILING DATE: 14-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 186,097
; FILING DATE: 25-APR-1988
; SEQ ID NO:1:
; LENGTH: 1886
5210183-1

Query Match 1.3%; Score 36.8; DB 6; Length 1886;
Best Local Similarity 45.5%; Pred. No. 0.66;
Matches 131; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 694 AACAGGAAAATTTATCAGAGATGAAAGAACTGAAGTACAAGAGCAGATGAATCTCCTAAA 753
Db 261 AGCGTGGAACTGGCCTAAAGAAATATAACGCGTTACTTAAGGAAAATGAGGAGCTCAAGGT 320
QY 754 TTCAAGATGGGAATGCCCTCAGGCTAGCTAGCATGGAAAAACAAAGCAATTTACATAGAGT 813
Db 321 AGAAGCTGAAAAATATCTATCTTATGCTGACGATAAAGAAAAAGATCCTCAATCTAGAGC 380
QY 814 TTTAATGGATCTCCAGAAATCAGAAACTGAAAGAGTTGAATGACTGGCTTAACAAAAACAGA 873
Db 381 ATTAATGGGTGAAAATCAAGATCTTCGAAAAAGAGAGGGACAAATATCAGGACAAAATAGA 440
QY 874 AGAAGAACAAAGGAAAATGGAGGAAGAGCGCTCTTGGACCTGATCTTGAAGACCTAAACG 933
Db 441 AGAAGCTGAAAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 500
QY 934 CCAAGTACAAACATAAAGGTGCTTCAAGAGATCTAGAACAAAGAAACA 981
Db 501 TCAATAGACGACAGATAGCATTTATCAAGAACAAACAAACAAACAAACATCA 548

RESULT 8
US-08-887-534A-22

; Sequence 22, Application US/08887534A
; Patent No. 6455323
; GENERAL INFORMATION:
; APPLICANT: Holden, David W.
; TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS
; NUMBER OF SEQUENCES: 106
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive/6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: United States of America
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/887,534A
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Rin-Laures, Li-Hsien
; REGISTRATION NUMBER: 33,547
; REFERENCE/DOCKET NUMBER: 28341/33996
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 474-6300
; TELEFAX: (312) 474-0448
; TELEX: (312) 474-6600
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1845 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
; DESCRIPTION: /desc = "DNA (genomic) (p14cl5)"
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 3..788
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 856..1842
; US-08-887-534A-22

Query Match 1.3%; Score 36.6; DB 4; Length 1845;
Best Local Similarity 48.0%; Pred. No. 0.74;
Matches 135; Conservative 0; Mismatches 144; Indels 2; Gaps 1;
QY 630 TGGATTGACAGCCCATCAGGCGCGGTTGGTAATATTCTACAATTTGGGAAGTAAGCTGA 689
Db 349 TGGCTTTGGATAAACATATGATGATGTAACAATTAGTTATCCCTACCTGACAATCAGC 408
QY 690 TTGGAACAGGAAAATTTATCAGAGATGAAGAAACTGAAGTACAAGAGCAGATGAATCTCC 749
Db 409 TACTTACCCAACTATTTACGGAAGTGAAGAAAGGTGCTGTAATGTTGATATGATTTTAC 468
QY 750 TAAATTCAGATGGGAATGCCCTCAGGCTAGCTAGCATGGAAAAACAAAGCAATTTACATA 809
Db 469 AAATCGTCAACTTGGATGGGCTACAACTATCCTTCACGATTAAAGATAGTATTTTCATC 528
QY 810 GAGTTTAAATGGATCTCCAGAAATCAGAAACTGAAAGAGTTGAATGACTGGCTTAACAAAA 869
Db 529 AAATTTCTATGATTTCTT--GAAACATTAAGAATCAATATGAAGCATTAGCTTATAAAAT 586
QY 870 CAGAAGAAAGAAAGGAAAATGGAGGAGAGCCCTCTTGA 910
Db 587 CAATGAGCATATGTCAAAATTTCAATTAATTGCTCAGGCA 627

RESULT 9
US-09-007-005-17


```
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; EARLIER FILING DATE: 1998-01-14
; EARLIER FILING DATE: 1997-01-27
; EARLIER FILING DATE: 1997-01-27
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17
```

```
Query Match 1.3%; Score 36.2; DB 4; Length 289;
Best Local Similarity 6.1%; Pred. No. 0.26;
Matches 14; Conservative 99; Mismatches 118; Indels 0; Gaps 0;

QY 762 GCGAATGCCTCAGGCTAGCATGGAAACAAAGCAATTTACATAGAGTTTAAATGG 821
Db 3 RGRGRARARURURARURURURARURURURARURURURARURURARURURURUR 62

QY 822 ATCTCCAGAAATCAGAACTGAAAGAGTTGAATGACTGGCTAAGCAAAACAGAAAGAA 881
Db 63 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 122

QY 882 CAAGGAAATGGAGGAGAGCCTCTTGGACCTGATCTTGAAGACCTAAACGCCAAGTAC 941
Db 123 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 182

QY 942 AACACATAAGGTGCTTCAAGAAGATCTAGAACAGCAAGCAAGTCAAGGTCA 992
Db 183 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 233
```

```
RESULT 10
US-09-244-796-17
; Sequence 17, Application US/09244796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; EARLIER FILING DATE: 1999-02-05
; EARLIER FILING DATE: 1997-01-27
; EARLIER FILING DATE: 1997-11-06
; EARLIER FILING DATE: 1997-11-06
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
```

```
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; NAME/KEY: misc_feature
; LOCATION: (1)...(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

Query Match 1.3%; Score 36.2; DB 4; Length 289;
Best Local Similarity 6.1%; Pred. No. 0.26;
Matches 14; Conservative 99; Mismatches 118; Indels 0; Gaps 0;

QY 762 GCGAATGCCTCAGGCTAGCATGGAAACAAAGCAATTTACATAGAGTTTAAATGG 821
Db 3 RGRGRARARURURARURURURARURURURARURURURARURURARURURURUR 62

QY 822 ATCTCCAGAAATCAGAACTGAAAGAGTTGAATGACTGGCTAAGCAAAACAGAAAGAA 881
Db 63 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 122

QY 882 CAAGGAAATGGAGGAGAGCCTCTTGGACCTGATCTTGAAGACCTAAACGCCAAGTAC 941
Db 123 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 182

QY 942 AACACATAAGGTGCTTCAAGAAGATCTAGAACAGCAAGCAAGTCAAGGTCA 992
Db 183 RNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNRNR 233
```

```
RESULT 11
US-09-014-969-14
; Sequence 14, Application US/09014969
; Patent No. 5965397
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Racie, Lisa A.
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/014,969
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2447 base pairs
; TYPE: nucleic acid
```


US-08-630-915A-193

RESULT 14
 US-08-139-937-12
 ; Sequence 12, Application US/08139937
 ; Patent No. 5821070
 ; GENERAL INFORMATION:
 ; APPLICANT: LEE, WEN-HWA
 ; APPLICANT: SHAN, BEI
 ; TITLE OF INVENTION: CELLULAR GENES ENCODING
 ; TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS
 ; NUMBER OF SEQUENCES: 14
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: CAMPBELL AND FLORES
 ; STREET: 4370 LA JOLLA VILLAGE DRIVE
 ; CITY: SAN DIEGO
 ; STATE: CALIFORNIA
 ; COUNTRY: USA
 ; ZIP: 92122
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ;

;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/139,937
;; FILING DATE: 20-OCT-1993
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/979,156
;; FILING DATE: 20-NOV-1992
;; ATTORNEY/AGENT INFORMATION:
;; NAME: CAMPBELL, CATHRYN
;; REGISTRATION NUMBER: 31,815
;; REFERENCE/DOCKET NUMBER: P-CJ 9370
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 619-535-9001
;; TELEFAX: 619-535-8949
;; INFORMATION FOR SEQ ID NO: 12:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4868 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: CDNA
US-08-139-937-12

Query Match 1.3%; Score 35.6; DB 1; Length 4868;
Best Local Similarity 51.9%; Pred. No. 3;
Matches 80; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 864 CAAAAACAGAGAAAGAACAGGAAATGGAGGAGCCTCTTGGACCTGATCTTGAAG 923
Db 1921 CAGAAAAAGAAATCTGACAAATGAAATACAAAAAGAGCAAGCGAATATCTGAATTAG 1980

QY 924 ACCTAAAACGCCAAGTACAAACATAGGTGCTTCAAGAAAGATCTAGAACAAAGCAAG 983
Db 1981 AAATAATAAATTCATCATTTGAAATATTTTGAAGAAAAAGAGCAAGCAAGTACAGA 2040

QY 984 TCAGGGTCAATTCTCTCACTCACATGGTGGTGGT 1017
Db 2041 TGAAGAAAAATCAAGCACTGCCATGGAGATGCT 2074

RESULT 15
PCT-US93-11310-12
;; Sequence 12, Application PC/TUS9311310
;; GENERAL INFORMATION:
;; APPLICANT: BOARD OF REGENTS OF THE UNIVERSITY OF TEXAS SYSTEM
;; TITLE OF INVENTION: CELLULAR GENES ENCODING
;; TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS
;; NUMBER OF SEQUENCES: 14
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: CAMPBELL AND FLORES
;; STREET: 4370 LA JOLLA VILLAGE DRIVE
;; CITY: SAN DIEGO
;; STATE: CALIFORNIA
;; COUNTRY: USA
;; ZIP: 92122
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: PatentIn Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: PCT/US93/11310
;; FILING DATE: 19-NOV-1993
;; CLASSIFICATION:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: CAMPBELL, CATHRYN
;; REGISTRATION NUMBER: 31,815
;; REFERENCE/DOCKET NUMBER: FP-CJ 9790
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 619-535-9001
;; TELEFAX: 619-535-8949
;; INFORMATION FOR SEQ ID NO: 12:

;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 4868 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: CDNA
PCT-US93-11310-12

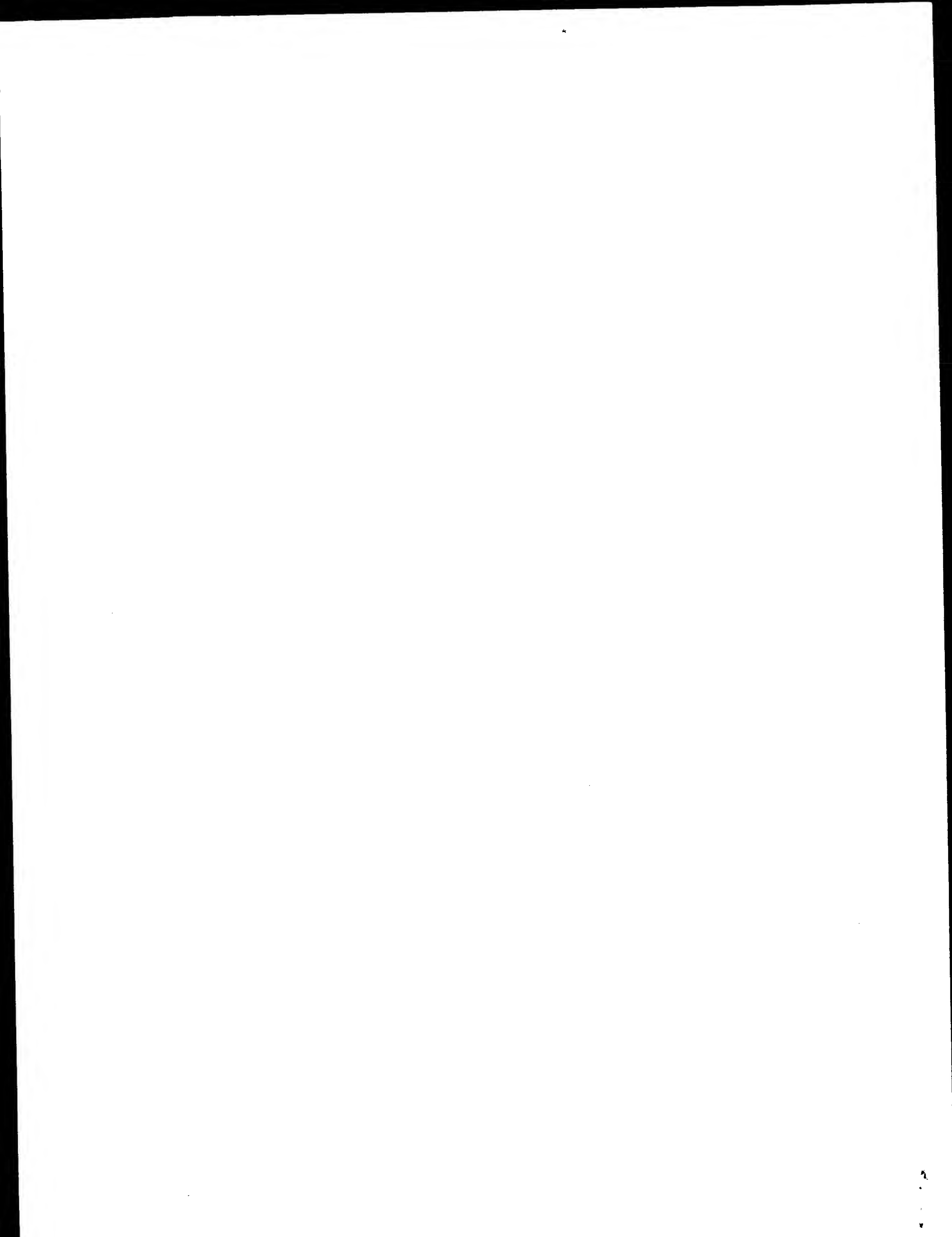
Query Match 1.3%; Score 35.6; DB 5; Length 4868;
Best Local Similarity 51.9%; Pred. No. 3;
Matches 80; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 864 CAAAAACAGAGAAAGAACAGGAAATGGAGGAGCCTCTTGGACCTGATCTTGAAG 923
Db 1921 CAGAAAAAGAAATCTGACAAATGAAATACAAAAAGAGCAAGCGAATATCTGAATTAG 1980

QY 924 ACCTAAAACGCCAAGTACAAACATAGGTGCTTCAAGAAAGATCTAGAACAAAGCAAG 983
Db 1981 AAATAATAAATTCATCATTTGAAATATTTTGAAGAAAAAGAGCAAGCAAGTACAGA 2040

QY 984 TCAGGGTCAATTCTCTCACTCACATGGTGGTGGT 1017
Db 2041 TGAAGAAAAATCAAGCACTGCCATGGAGATGCT 2074

Search completed: April 25, 2003, 17:43:39
Job time : 98.8837 secs



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 17:39:49 ; Search time 183.448 Seconds
(without alignments)
16376.941 Million cell updates/sec

Title: US-09-845-416-9_COPY_540_3300
Perfect score: 2761
Sequence: 1 gccagacattattgactgga.....aagtgcaagttaacacagga 2761

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 709820 seqs, 544064369 residues

Total number of hits satisfying chosen parameters: 1410746

Minimum DB seq length: 0
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	190.2	6.9	256	10	US-09-864-761-21956
2	178	6.4	466	10	US-09-864-761-21956
3	153.4	5.6	467	10	US-09-864-761-6092
4	151	5.5	151	10	US-09-864-761-11083
5	46.4	1.7	425	9	US-09-864-761-27715
6	43	1.6	436	10	US-09-918-995-35971
7	42.2	1.5	423	10	US-09-960-352-10742
8	41.8	1.5	440	9	US-09-864-761-18355
9	41.8	1.5	440	9	US-10-184-644-442
10	41.6	1.5	557	10	US-10-184-634-442
11	40.6	1.5	425	10	US-09-864-761-7315
12	40.4	1.5	428	10	US-09-960-352-4010
13	39.8	1.4	418	10	US-09-864-864-204
14	39.6	1.4	1381	9	US-09-960-352-12395
15	39.6	1.4	1594	9	US-09-822-846-343
16	39.6	1.4	1594	9	US-10-174-590-183
17	39.6	1.4	1594	9	US-10-176-758-183
18	39.6	1.4	1594	9	US-10-175-737-183
19	39.6	1.4	1594	9	US-10-173-706-183
					US-10-175-738-183

C 20	39.6	1.4	1594	9	US-10-175-752-183	Sequence 183, App
C 21	39.6	1.4	1594	9	US-10-176-482-183	Sequence 183, App
C 22	39.6	1.4	1594	9	US-10-176-757-183	Sequence 183, App
C 23	39.6	1.4	1594	9	US-10-176-913-183	Sequence 183, App
C 24	39.6	1.4	1594	9	US-10-180-552-183	Sequence 183, App
C 25	39.6	1.4	1594	9	US-10-180-557-183	Sequence 183, App
C 26	39.6	1.4	1594	9	US-10-173-700-183	Sequence 183, App
C 27	39.6	1.4	1594	9	US-10-174-572-183	Sequence 183, App
C 28	39.6	1.4	1594	9	US-10-174-579-183	Sequence 183, App
C 29	39.6	1.4	1594	9	US-10-174-582-183	Sequence 183, App
C 30	39.6	1.4	1594	9	US-10-174-588-183	Sequence 183, App
C 31	39.6	1.4	1594	9	US-10-175-739-183	Sequence 183, App
C 32	39.6	1.4	1594	9	US-10-175-740-183	Sequence 183, App
C 33	39.6	1.4	1594	9	US-10-175-743-183	Sequence 183, App
C 34	39.6	1.4	1594	9	US-10-176-488-183	Sequence 183, App
C 35	39.6	1.4	1594	9	US-10-176-492-183	Sequence 183, App
C 36	39.6	1.4	1594	9	US-10-176-747-183	Sequence 183, App
C 37	39.6	1.4	1594	9	US-10-176-750-183	Sequence 183, App
C 38	39.6	1.4	1594	9	US-10-176-985-183	Sequence 183, App
C 39	39.6	1.4	1594	9	US-10-176-987-183	Sequence 183, App
C 40	39.6	1.4	1594	9	US-10-176-991-183	Sequence 183, App
C 41	39.6	1.4	1594	9	US-10-176-992-183	Sequence 183, App
C 42	39.6	1.4	1594	9	US-10-176-993-183	Sequence 183, App
C 43	39.6	1.4	1594	9	US-10-184-658-183	Sequence 183, App
C 44	39.6	1.4	1594	9	US-10-173-695-183	Sequence 183, App
C 45	39.6	1.4	1594	9	US-10-173-697-183	Sequence 183, App

ALIGNMENTS

RESULT 1
US-09-864-761-21956
; Sequence 21956, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30

Mon Apr 28 09:28:40 2003

PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 6092
LENGTH: 466
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC004468.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
US-09-864-761-6092

Query Match 6.9%; Score 190.2; DB 10; Length 256;
Best Local Similarity 98.5%; Pred. No. 1.4e-47;
Matches 192; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 431 CATTTGGAAGCTCCTGAACACAGTCATTGGCAGTTTCATTGATGGAGAGTGAAGTAAAC 490
Db 1 CATTTGGAAGCTCCTGAACACAGTCATTGGCAGTTTCATTGATGGAGAGTGAAGTAAAC 60
QY 491 CTGGACCGTTATCAACAGCTTTAGAGAAGTATTATCGTGGCTTCTTCTGCTGAGGAC 550
Db 61 CTGGACCGTTATCAACAGCTTTAGAGAAGTATTATCGTGGCTTCTTCTGCTGAGGAC 120
QY 551 ACATTGCAAGCACAAAGGAGAGATTCTTAATGATGTGGAAGTGGTGAAGACCAAGTTTCAT 610
Db 121 ACATTGCAAGCACAAAGGAGAGATTCTTAATGATGTGGAAGTGGTGAAGACCAAGTTTCAT 180

QY 611 ACTCATGAGGGGTAC 625
Db 181 ACTCATGAGGGTAAAC 195

RESULT 2
US-09-864-761-6092
; Sequence 6092, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeonica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27

Query Match 6.4%; Score 178; DB 10; Length 466;
Best Local Similarity 97.3%; Pred. No. 1.1e-43;
Matches 181; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 419 TTTCTTTCACAGCATTGGAAGCTCCTGAAGACAAAGTCATTGGCAGTTTCATTGATGGAG 478
Db 281 TTTATTGTGCAGCATTGGAAGCTCCTGAAGACAAAGTCATTGGCAGTTTCATTGATGGAG 340
QY 479 AGTGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAGAAGTATTATCGTGGCTTCTT 538
Db 341 AGTGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAGAAGTATTATCGTGGCTTCTT 400
QY 539 TCTGCTGAGGACACATTGCAAGCACAAAGGAGAGATTCTTAATGATGTGGAAGTGGTGA 598
Db 401 TCTGCTGAGGACACATTGCAAGCACAAAGGAGAGATTCTTAATGATGTGGAAGTGGTGA 460

QY 599 GACCAG 604
Db 461 GACCAG 466

RESULT 3
US-09-864-761-11083
; Sequence 11083, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.

TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 46-LIB3058-032-Q1-K1-D10
US-09-960-352-10742

Query Match 1.6%; Score 43; DB 10; Length 436;
Best Local Similarity 48.6%; Pred. No. 0.019;
Matches 118; Conservative 0; Mismatches 125; Indels 0; Gaps 0;

QY 694 AACAGGAAATATCAGAGATGAAGAACTGAAGTACAGAGAGAGATGAAATCTCTCTAAA 753
Db 431 AATAGAAAAAGAAAGAAAAACAGACAATAAAATAAAAAATAAAAAATAAAAAATA 372
QY 754 TTCAAGATGGGAATGCCTCAGGCTACCTAGCATGGAAAAACAAAGCAATTTACATAGAT 813
Db 371 TAAAAAGAAATAAAACAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAATAAT 312
QY 814 TTTAATGGATCTCCAGAAATCAGAAATGAAAGAGTTGAATGACTGGCTAACAACACAGA 873
Db 311 AAGAAGACAAAAACAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATA 252
QY 874 AGAAAGAACAAAGGAAAAATGGAGGAAGAGCCCTCTTGGACCTGATCTTGAAGACCTAAACG 933
Db 251 AAAAAAGAAAGAAAAATAAAAAATAAAAAATAAAAAATAAAAAATAAAAAATA 192
QY 934 CCA 936
Db 191 GCA 189

RESULT 7
US-09-864-761-18355
Sequence 18355, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Acomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670

QY 802 TTTACATAGAGTTTAAATGGATCTCCAGAATCAGAACTGAAAGAGTTGAATGACTGGCT 861
Db 1 TTTACATAGAGTTTAAATGGATCTCCAGAATCAGAACTGAAAGAGTTGAATGACTGGCT 60
QY 862 AACAAAAACAGAAAGAAAGAAAGAAATGGAGGAGAGCTCTTGGACCTGATCTTGA 921
Db 61 AACAAAAACAGAAAGAAAGAAAGAAATGGAGGAGAGCTCTTGGACCTGATCTTGA 120
QY 922 AGACCTAAACGCCCAAGTACAAACACATAAG 952
Db 121 AGACCTAAACGCCCAAGTACAAACACATAAG 151

RESULT 5
US-09-918-995-35971
Sequence 35971, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 35971
LENGTH: 425
TYPE: DNA
ORGANISM: Homo sapiens
US-09-918-995-35971

Query Match 1.7%; Score 46.4; DB 9; Length 425;
Best Local Similarity 57.1%; Pred. No. 0.0017;
Matches 105; Conservative 0; Mismatches 76; Indels 3; Gaps 1;
QY 2 CCAGACCTATTGACTGGAATAGTGTGGTTGCCAGCAGTCAGCCACACACGACTGGAA 61
Db 220 CCAGACCTAATTGACATGAAGAGTGTG---AAGCATAGATCCCAACAAAGACAATCTGAGA 276
QY 62 CATGCATTCAACATCGCCAGATATCAATTAGGCATAGAGAACTACTCGATCCTGAAGAT 121
Db 277 GAGGCTTCAGAATTGCAGAACAAAGAAATTAATAATCCCAAGATTCCTGGACCAAGAT 336
QY 122 GTTGATACCACTATCCAGATAAGAGTCAATCTTAATGTACATCACATCCTCTTCAA 181
Db 337 GTGGATGTTGTGATCCTGATGATGAAGAAAGTCCATCATGACCTATGTGGCACAGTTTCTGCAG 396
QY 182 GTTT 185
Db 397 TATT 400

RESULT 6
US-09-960-352-10742/C
Sequence 10742, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 10742
LENGTH: 436

Mon Apr 28 09:28:40 2003

PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 7315
LENGTH: 557
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC008739.5
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.2
US-09-864-761-7315

Query Match 1.5%; Score 41.6; DB 10; Length 557;
Best Local Similarity 53.0%; Pred. No. 0.059;
Matches 89; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
QY 159 TGTACATCACATCCTCTTCCAAAGTTTGCCTCAACAAGTGAGCATTGAAGCCATCCAGG 218
DB 216 TATGCAAGCATATAAATTACCAAAATATTTCTCAAAAATGAGAAATGAAACCTTTAGGG 275
QY 219 AACTGGAATGTTGCGCAAGGCCACTAAAGTGACTAAAGAAGAACATTTTCAGTTACATC 278
DB 276 TATACTAGTAATTTGTACTAAAGTTATTTCTACCCCAAGAGACCCAGGTTTCTGTAGTTC 335
QY 279 ATCAATGCACTATTTCTCAACAGATACCGGTCAGTCTAGCACAGGGAT 326
DB 336 TCACATCACATCTTTCTATACAAATCCTGCTGAGCAGTGTCCAGGCAT 383

RESULT 11
US-09-960-352-4010/C
Sequence 4010, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathialagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 4010
LENGTH: 425
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 18-BOVMS1-007-Q1-E1-E5
US-09-960-352-4010

Query Match 1.5%; Score 40.6; DB 10; Length 425;
Best Local Similarity 46.3%; Pred. No. 0.099;
Matches 133; Conservative 0; Mismatches 154; Indels 0; Gaps 0;
QY 700 AAAATTATCAGAAGATGAAGAACTGAAGTACAGCAGATGAATCTCTAAATCAAG 759
DB 424 AAAAAAAAAAAAAAAAAACAAAAAATAAAAAAAAAAAAAAAAAAAAAA 365
QY 760 ATGGGAATGCCTCAGGTAGCTAGCATGGAACAAAGCAATTTACATAGAGTTTAAT 819

QY 2135 GACCTGAACACACAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAAGCAGCTG 2194
DB 429 GDS.WB..TACTYCM.KGKBYSWSCNN...A.CBBNGKBNRBHSY...TTY.NSM.H.C. 370
QY 2195 CATGAAGCCACAGGGACTTTGGTCCAGCATCTCAGCAGCTTTCTTTCCAGCTGTGTCAG 2254
DB 369 T...CYKHK..NHRCDWCY.WC.YYCASNKW...K.YN.A..WC.HTB.A..SB.MCMMR 310
QY 2255 GGTCCCTGGGAGAGGCCATCTCGCCCAACAAAGTGCCCTACTATATCAACACGAGACT 2314
DB 309 R.YS..T.M.RBAKD.BAS.MNR..SDMC.T.W.SBB.CYHMB...HN.S.RMN...BMH 250
QY 2315 CAACAACACTTGTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTG 2374
DB 249 KW..HAKCDHS...BN.YH..CY...WMAKRM.BAHBTM.THMY..ST.CRDDC.BH.Y. 190
QY 2375 AATAATGTCAGATCTCAGCTTATAGGACTGCCATGAACTCCGAAGACTGCAGAGGCC 2434
DB 189 DR.TT.HTYCSM.SH.THSBTNT.N.HA.BNMR...GRA.YBC...B.GC.WRRKRBMVMT 130
QY 2435 CTTTGTGATCTCTTGCAGCCTGTGAGCTGCTGATGCTGATGCTGCTGAGCAGCACAACCTC 2494
DB 129 MA.ABW...SCNWW..K.SHMNB.C.H.SMNAA.HWKRNC.K.DB.GDYCAK.ARRN.RS. 70
QY 2495 AAGCAAAATGACGAGCCCA 2513
DB 69 S..H.A.RHBMBCYSB..M 51

RESULT 10
US-09-864-761-7315
Sequence 7315, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aecomica-x-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670

Search completed: April 25, 2003, 23:58:14
Job time : 192.615 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 07:56:07 ; Search time 2252.17 Seconds
(without alignments)
17070.103 Million cell updates/sec

Title: US-09-845-416-10_COPY_1800_3120
Perfect score: 1321
Sequence: 1 cgactttccagcagttcaga.....ctaataataagccagagatc 1321

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 3813070

Minimum DB seq length: 0
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rod:*
- 36: em_htg_mam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1105.2	83.7	4402	6	E30219	E30219 Shortened d
2	906	68.6	4402	6	E30220	E30220 Shortened d
3	770.4	58.3	4075	6	E30221	E30221 Shortened d
4	766.2	58.0	3747	6	E30218	E30218 Shortened d
5	737.6	55.8	2654	5	FSCDYSTRO	M37645 Torpedo cal
6	712.6	53.9	3163	6	E30223	E30223 Shortened d
7	619.2	46.9	3521	5	AF339031	AF339031 Danio rer
8	604	45.7	2110	9	HUMDMDX	M92650 Human Duche
9	577.8	43.7	4658	9	BC028720	BC028720 Homo sapi
10	555	42.0	3499	9	HSU43519	X83506 M.musculus
11	551.8	41.8	3073	10	AF195788	U43519 Human dyst
12	551.8	41.8	3140	10	AF195787	AF195788 Rattus no
13	544	41.2	1240	10	RNAPDY3	AF195787 Rattus no
14	493.8	37.4	3172	6	E30222	X69767 R.norvegicu
15	410.4	31.1	1961	5	XLDYSTROP	E30222 Shortened d
16	379.4	28.7	1195	5	AF339032	X99700 X.laavis mR
17	375	28.4	1993	5	SCDYSTROP	AF339032 Danio rer
18	369.8	28.0	3650	3	SPAJ3356	X99702 S.canicula
19	286.4	21.7	1482	5	XLUTROPHI	AJ223356 Strongylo
20	285.4	21.6	1428	5	SCU43517	X99701 X.laavis mR
21	273.4	20.7	500	9	HUMDYSTR15	U43517 Scylliorhinu
22	272.6	20.6	1737	5	SCUTROPHI	L05649 Homo sapien
23	265	20.1	506	10	AB011666	X99703 S.canicula
24	241.8	18.3	1440	5	GSU43518	AB011666 Rattus no
25	241.6	18.3	384	10	RNADYI	U43518 Gobius sp.
26	240	18.2	1428	10	MMU43520	X65468 R.norvegicu
27	225	17.0	1431	3	PSDYSTROP	U43520 Mus musculu
28	221.6	16.8	1413	3	ASDYSTROP	X99738 Pectinidae
29	208.6	15.8	1338	3	BLDYSTROP	X99736 B.lanceolat
30	203	15.4	351	9	HUMDYSTO6	M86889 H.sapiens d
31	191.8	14.5	1729	3	CIDYSTRO	X99756 C.intestina
32	179.8	13.6	374	10	S62620	S62620 (mdx3Cv)=dy
33	161.2	12.2	433	9	HUMDYSTO8	M86891 H.sapiens d
34	157	11.9	300	9	S38776	S38776 Homo sapien
35	153.2	11.6	2962	3	DMDYDTRO	X99757 D.melanogas
36	151	11.4	332	9	HUMDYSTO1	M86884 H.sapiens d
37	149.4	11.3	989	9	HSDYSTX60	Z11860 H.sapiens g
38	149	11.3	1470	9	HSDYSRP	X15488 Human mRNA
39	122	9.2	360	9	HUMDYSTR14	L05648 Homo sapien
40	94	7.1	250	9	HUMDYSTO7	M86890 H.sapiens d
41	83.8	6.3	219	9	HUMDYSTO5	M86888 H.sapiens d
42	83.8	6.3	343	9	AF213441	AF213441 Homo sapi
43	82.2	6.2	200	10	S38777	S38777 Rattus norv
44	82	6.2	167	9	HUMDYSTO2	M86885 H.sapiens d
45						

ALIGNMENTS

RESULT 1
E30219
LOCUS
DEFINITION Shortened dystrophin.
ACCESSION E30219
VERSION E30219.1 GI:13017026
KEYWORDS JP 1999318467-A/2.
SOURCE unidentified.
ORGANISM unidentified.
REFERENCE 1 (bases 1 to 4402)
AUTHORS Sinichi, T.
TITLE Shortened dystrophin
JOURNAL Patent: JP 1999318467-A 2 24-NOV-1999;
SCIENCE & TECH AGENCY, NATIONAL CENTER OF NEUROLOGY AND PSYCHIATRY

4402 bp DNA linear PAT 18-JUN-2001

QY	476	GTCAATGACCTTGCTCGCCAGCTTACCACTTTGGCATTTCAGCTCTCACCGTATAACCTC	535
Db	2062	GTCAATGACCTTGCTCGCCAGCTTACCACTTTGGCATTTCAGCTCTCACCGTATAACCTC	2121
QY	536	AGCACTCTGGAAGACCTGAACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGA	595
Db	2122	AGCACTCTGGAAGACCTGAACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGA	2181
QY	596	GTCAGGCAGCTGCATGAAGCCCAACAGGGACTTTGGTCCAGCATCTCAGCACTTCTCTTTCC	655
Db	2182	GTCAGGCAGCTGCATGAAGCCCAACAGGGACTTTGGTCCAGCATCTCAGCACTTCTCTTTCC	2241
QY	656	ACGTCTGTCCAGGTCCTTGGGAGAGAGCCATCTCGSCAAACAAAGTGCCCTACTATATC	715
Db	2242	ACGTCTGTCCAGGTCCTTGGGAGAGAGCCATCTCGSCAAACAAAGTGCCCTACTATATC	2301
QY	716	AACCACGAGACTCAAAACAACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCACTCT	775
Db	2302	AACCACGAGACTCAAAACAACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCACTCT	2361
QY	776	TTAGCTGACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGA	835
Db	2362	TTAGCTGACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGA	2421
QY	836	CTGCAGAAAGGCCCTTTTGCTTGGATCTCTTGAGCCTGTTCAGCTGCATGTGATGCCCTTGGAC	895
Db	2422	CTGCAGAAAGGCCCTTTTGCTTGGATCTCTTGAGCCTGTTCAGCTGCATGTGATGCCCTTGGAC	2481
QY	896	CAGCACAACTCAAGCAAAATGACCAAGCCCATGGATATCCTGCAGATTATTAATGTTTG	955
Db	2482	CAGCACAACTCAAGCAAAATGACCAAGCCCATGGATATCCTGCAGATTATTAATGTTTG	2541
QY	956	ACCACTATTTATGACCGCCTGGAGCAAGAGCACAACTTTGGTCAACGTCCCTCTCTGC	1015
Db	2542	ACCACTATTTATGACCGCCTGGAGCAAGAGCACAACTTTGGTCAACGTCCCTCTCTGC	2601
QY	1016	GTGGATATGTGCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGATC	1075
Db	2602	GTGGATATGTGCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGATC	2661
QY	1076	CGTGCTCTGCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAG	1135
Db	2662	CGTGCTCTGCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAG	2721
QY	1136	TACAGATACCTTTTCAAGCAAGTGGCAAGTCAACAGGATTTTGTACCAAGCGCAGGCTG	1195
Db	2722	TACAGATACCTTTTCAAGCAAGTGGCAAGTCAACAGGATTTTGTACCAAGCGCAGGCTG	2781
QY	1196	GGCTCCTTCTGCATGATTCTATCCAAATTTCCAAGACAGTTGGGTGAAGTTGCATCCTTT	1255
Db	2782	GGCTCCTTCTGCATGATTCTATCCAAATTTCCAAGACAGTTGGGTGAAGTTGCATCCTTT	2841
QY	1256	GGGGCAGTAACATTGAGCCAAGTGTCCGGAGCTGCTTCCAAATTTGCTAATAATAAGCCA	1315
Db	2842	GGGGCAGTAACATTGAGCCAAGTGTCCGGAGCTGCTTCCAAATTTGCTAATAATAAGCCA	2901
QY	1316	GAGATC	1321
Db	2902	GAGATC	2907

RESULT 3	4075 bp	DNA	linear	PAT 18-JUN-2001
E30221				
LOCUS	E30221			
DEFINITION	Shortened dystrophin.			
ACCESSION	E30221			
VERSION	E30221.1			
KEYWORDS	GI:13017028			
SOURCE	JP 1999318467-A/4.			
ORGANISM	unidentified.			
	unidentified			
	unclassified.			
REFERENCE	1 (bases 1 to 4075)			
AUTHORS	Sinichi, T.			

TITLE	Shortened dystrophin
JOURNAL	Patent: JP 1999318467-A 4 24-NOV-1999; SCIENCE & TECH AGENCY, NATIONAL CENTER OF NEUROLOGY AND PSYCHIATRY
COMMENT	OS Unidentified PN JP 1999318467-A/4 PD 24-NOV-1999 PF 08-MAY-1998 JP 1998142134 PR PI SINICHI TAKEDA PC C12N15/09,A61K48/00,C12N15/00 CC Strandedness: Both; CC Topology: Linear; FH Key FT source 1..4075 FT Location/Qualifiers Location/Qualifiers 1..4075 /organism='Unidentified'. /organism="unidentified" /db_xref="taxon:32644"
FEATURES	
source	
BASE COUNT	1237 a 910 c 936 g 992 t
ORIGIN	

Query Match		58.3%;	Score 770.4;	DB 6;	Length 4075;
Best Local Similarity		99.2%;	Pred. No. 2.6e-213;		
Matches 774; Conservative		0;	Mismatches 6;	Indels	0; Gaps
QY	542	CTGGAAGACCTGAACACACCAGATGGAAGCTTCTGCAGGTGGCGTTCGAGGACCGAGTCAGG	601		
Db	1801	CTTAAGGTATTGAACACACCAGATGGAAGCTTCTGCAGGTGGCGTTCGAGGACCGAGTCAGG	1860		
QY	602	CAGCTGCATGAAGCCCAACAGGGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGTCT	661		
Db	1861	CAGCTGCATGAAGCCCAACAGGGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGTCT	1920		
QY	662	GTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCCAC	721		
Db	1921	GTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCCAC	1980		
QY	722	GAGACTCAAAACAACTTGCTGGGACCATCCCAAATGACAGAGCTCTACCAGTCTTTAGCT	781		
Db	1981	GAGACTCAAAACAACTTGCTGGGACCATCCCAAATGACAGAGCTCTACCAGTCTTTAGCT	2040		
QY	782	GACCTGAATAATGTGAGTCTCTCAGCTTATAGGACTGCCATGAACTCCGAAGACTGCAG	841		
Db	2041	GACCTGAATAATGTGAGTCTCTCAGCTTATAGGACTGCCATGAACTCCGAAGACTGCAG	2100		
QY	842	AAGGCCCTTTGCTTGGATCTCTTGAGCCTGTGAGCTGCATGTGATGCCCTTGGACCAAC	901		
Db	2101	AAGGCCCTTTGCTTGGATCTCTTGAGCCTGTGAGCTGCATGTGATGCCCTTGGACCAAC	2160		
QY	902	AACCTCAAGCAAAATGACCAAGCCATGGATATCCTGCAGATTATTAATTGTTTGACCACT	961		
Db	2161	AACCTCAAGCAAAATGACCAAGCCATGGATATCCTGCAGATTATTAATTGTTTGACCACT	2220		
QY	962	ATTTATGACCGCCTGGAGCAAGAGACACAACAATTTGGTCAACGTCCCTCTCTCGGTGGAT	1021		
Db	2221	ATTTATGACCGCCTGGAGCAAGAGACACAACAATTTGGTCAACGTCCCTCTCTCGGTGGAT	2280		
QY	1022	ATGTGCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCGTGTC	1081		
Db	2281	ATGTGCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCGTGTC	2340		
QY	1082	CTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAGACACATTTGGAAGACAAGTACAGA	1141		
Db	2341	CTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAGACACATTTGGAAGACAAGTACAGA	2400		
QY	1142	TACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAGCGCAGGCTGGCCCTC	1201		
Db	2401	TACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAGCGCAGGCTGGCCCTC	2460		
QY	1202	CTTCTGCATGATTTCTATCCAAATTCGAAGACAGTTTGGGTGAAGTTGCATCCTTTGGGGC	1261		
Db	2461	CTTCTGCATGATTTCTATCCAAATTCGAAGACAGTTTGGGTGAAGTTGCATCCTTTGGGGC	2520		

||||| 1895 TTTATGACCGCTGGAGCAAGAGCACAAATTTGGTCAACGTCCTCTCTCGGTGGATA 1954
QY 1023 TGTGCTGAACCTGGCTGCTGAATGTTTATGATACGGACGAACAGGAGGATCCGTGTCC 1082
Db 1955 TGTGCTGAACCTGGCTGCTGAATGTTTATGATACGGAGCAACAGGAGGATCCGTGTCC 2014
QY 1083 TGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAGAT 1142
Db 2015 TGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAGAT 2074
QY 1143 ACCTTTTCAAGCAAGTGGCAAGTCAACAGGATTTTGTGACGACGCGAGGCTGGCCCTCC 1202
Db 2075 ACCTTTTCAAGCAAGTGGCAAGTCAACAGGATTTTGTGACGACGCGAGGCTGGCCCTCC 2134
QY 1203 TTCTGTCATGATTTCTATCCAAATTTCCAAAGACAGTTGGGTGAAGTTGCATCCTTTGGGGGCA 1262
Db 2135 TTCTGTCATGATTTCTATCCAAATTTCCAAAGACAGTTGGGTGAAGTTGCATCCTTTGGGGGCA 2194
QY 1263 GTAAACATTGAGCCCAAGTGTCCGGAGCTGCTTCCAAATTTGCTAATAATAAGCCAGAGATC 1321
Db 2195 GTAAACATTGAGCCCAAGTGTCCGGAGCTGCTTCCAAATTTGCTAATAATAAGCCAGAGATC 2253

RESULT 5
FSCDYSTRO 2654 bp mRNA linear VRT 08-SEP-1993
LOCUS Torpedo californica dystrophin mRNA, 3' end.
DEFINITION M37645
ACCESSION M37645.1 GI:397970
VERSION
KEYWORDS dystrophin.
SOURCE Torpedo californica, CDNA to mRNA.
ORGANISM Torpedo californica

REFERENCE
AUTHORS Yeadon, J.E., Lin, H., Dyer, S.M. and Burden, S.J.
TITLE Dystrophin is a component of the subsynaptic membrane
JOURNAL J. Cell Biol. 115 (4), 1069-1076 (1991)
MEDLINE 92064638
PUBMED 1720119

REFERENCE
AUTHORS Ravin, A.J., Dyer, S.M., Yeadon, J.E. and Burden, S.J.
TITLE Multiple dystrophin isoforms are associated with the postsynaptic membrane of Torpedo electric organ
JOURNAL J. Physiol. (Paris) 85 (3), 131-133 (1991)
MEDLINE 92291902
PUBMED 1818109
COMMENT On Sep 8, 1993 this sequence version replaced gi:213232.
Draft entry and computer-readable sequence for [Unpublished (1990)] kindly submitted by S.J. Burden, 06-AUG-1990.
Biology Dept, 16-820
MIT
Cambridge, MA 02139.

FEATURES
source 1. .2654
/organism="Torpedo californica"
/db_xref="taxon:7787"
1. .2654
/gene="dystrophin"
<1. .2654
/gene="dystrophin"
/codon_start=3
/product="dystrophin"
/protein_id="AAC38002.1"
/db_xref="GI:397971"
/translation="SGEQWKRLQISLQDFLTWNLKNDLRRQMP IGGDAPTVCQOND
VHRIFKRELKAKEPVMSALDTVHLFLADPAIRAPESLITGPKEIPEENIQNVAKRI
RKYAEVKVWDKLSNRSDWQKRIDEALKRLLEQDSMDLNLKLRQAEAIKDTWOP
VGDLIDSLQDHIKVKVFAEIAIPMKENVTHMNDLASQFTPPDIQLSPYNLQLEDL
NTRWKLQVSIIDELLKQLHEAHRDFGPTSQHFLSTSVQGPWERAISPKNKVPYYINHQT

QY 1262 AGTAACATTGAGCAAGTGTCCGGAGCTGCTTCCAAATTTGCTAATAATAAGCCAGAGATC 1321
Db 2521 AGTAACATTGAGCAAGTGTCCGGAGCTGCTTCCAAATTTGCTAATAATAAGCCAGAGATC 2580
RESULT 4
E30218 Shortened dystrophin.
LOCUS E30218
DEFINITION E30218
ACCESSION E30218
VERSION E30218.1 GI:13017025
KEYWORDS JP 1999318467-A/1.
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 3747)
AUTHORS Sinichi, T.
TITLE Shortened dystrophin
JOURNAL Patent: JP 1999318467-A 1 24-NOV-1999;
SCIENCE & TECH AGENCY, NATIONAL CENTER OF NEUROLOGY AND PSYCHIATRY
OS Unidentified
PN JP 1999318467-A/1
PD 24-NOV-1999
PF 08-MAY-1998 JP 1998142134
PR SINICHI TAKEDA
PI C12N15/09, A61K48/00, C12N15/00
PC C12N15/09, A61K48/00, C12N15/00
CC Strandedness: Both;
CC Topology: Linear;
FH Key
FT source 1. .3747
FT Location/Qualifiers
/organism="Unidentified".

FEATURES
Source 1. .3747
/organism="unidentified"
/db_xref="taxon:32644" 923 t

BASE COUNT 1112 a 853 c 859 g 923 t
ORIGIN

Query Match 58.0%; Score 766.2; DB 6; Length 3747;
Best Local Similarity 99.0%; Pred. No. 4.4e-212;
Matches 771; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 543 TGAAGACCTGAACACACAGATGGAAGTCTGCGAGGTCGCGGAGCCGAGTCAGGC 602
Db 1475 TGAATCTCTAAATTTCAAGATGGAAGCTTCTGCGAGGTCGCGGAGCCGAGTCAGGC 1534
QY 603 AGCTGCATGAAGCCACACAGGACTTTGGTCCAGCATCTCAGCAGCTTTCTTCCACGCTCTG 662
Db 1535 AGCTGCATGAAGCCACACAGGACTTTGGTCCAGCATCTCAGCAGCTTTCTTCCACGCTCTG 1594
QY 663 TCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTCCCTACTATATCAACCCAGC 722
Db 1595 TCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTCCCTACTATATCAACCCAGC 1654
QY 723 AGACTCAAAACAACTTGTCTGGGACCATCCCAAAATGACAGAGCTCTACCCAGTCTTTAGCTG 782
Db 1655 AGACTCAAAACAACTTGTCTGGGACCATCCCAAAATGACAGAGCTCTACCCAGTCTTTAGCTG 1714
QY 783 ACCTGAATAATGTGAGATCTCTGAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGA 842
Db 1715 ACCTGAATAATGTGAGATCTCTGAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGA 1774
QY 843 AGGCCCTTTGCTTGGATCTCTGAGCCTGTGAGCTGCATGATGCCTTGGACCAGCACA 902
Db 1775 AGGCCCTTTGCTTGGATCTCTGAGCCTGTGAGCTGCATGATGCCTTGGACCAGCACA 1834
QY 903 ACCTCAAGCAAAATGACACGCCCAGGATATCTCTGAGATTTATTAATTGTTGACCACCTA 962
Db 1835 ACCTCAAGCAAAATGACACGCCCAGGATATCTCTGAGATTTATTAATTGTTGACCACCTA 1894
QY 963 TTTATGACCGCCTGGAGCAAGAGCACAAATTTGGTCAACGTCCTCTCTCGGTGGATA 1022

Db	942	AGCCATCCAGGAAGTGGAGCCACAGGGACTTTGGTCCAGCATCTCAGCACATTTCTTTC	1001
QY	655	CACGTCTGTCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTCCCTACTATAT	714
Db	1002	CACGTCTGTCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTCCCTACTATAT	1061
QY	715	CAACCACGAGACTCAACAACTTGTCTGGGAGAGAGCCATCCCAAAATGACAGAGCTTACCAGTC	774
Db	1062	CAACCACGAGACTCAACAACTTGTCTGGGAGAGAGCCATCCCAAAATGACAGAGCTTACCAGTC	1121
QY	775	TTTAGCTGACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAACTCCGAAG	834
Db	1122	TTTAGCTGACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAACTCCGAAG	1181
QY	835	ACTGCAGAGAGCCCTTTGCTTGGATCTCTTGAGCCCTGTCAGCTGCATGTGATGCTTGA	894
Db	1182	ACTGCAGAGAGCCCTTTGCTTGGATCTCTTGAGCCCTGTCAGCTGCATGTGATGCTTGA	1241
QY	895	CCAGCACAACTCAAGCAAAATGACCCAGCCCATGGATATCTCTGCAGATTATTAATGTTT	954
Db	1242	CCAGCACAACTCAAGCAAAATGACCCAGCCCATGGATATCTCTGCAGATTATTAATGTTT	1301
QY	955	GACCACTATTATGACCGCTGGAGCAAGAGCACAACTTGGTCAACGTCCCTCTCTG	1014
Db	1302	GACCACTATTATGACCGCTGGAGCAAGAGCACAACTTGGTCAACGTCCCTCTCTG	1361
QY	1015	CGTGGATATGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGAGCAACAGGAGAT	1074
Db	1362	CGTGGATATGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGAGCAACAGGAGAT	1421
QY	1075	CCGTGCTCTGCTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAA	1134
Db	1422	CCGTGCTCTGCTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAA	1481
QY	1135	GTACAGATACCTTTTCAAGCAAGTGGCAAGTTCACACAGGATTTTGTGACCGCAGGCT	1194
Db	1482	GTACAGATACCTTTTCAAGCAAGTGGCAAGTTCACACAGGATTTTGTGACCGCAGGCT	1541
QY	1195	GGGCTCTCTGTCATGATTTATCCAAATTCACAGACAGTGGGTGAAGTTGCATCCTT	1254
Db	1542	GGGCTCTCTGTCATGATTTATCCAAATTCACAGACAGTGGGTGAAGTTGCATCCTT	1601
QY	1255	TGGGGCAGTAACATTGAGCCAAAGTGTCCGGAGCTGCTTCCAATTGCTAATAAGCC	1314
Db	1602	TGGGGCAGTAACATTGAGCCAAAGTGTCCGGAGCTGCTTCCAATTGCTAATAAGCC	1661
QY	1315	AGAGATC 1321	
Db	1662	AGAGATC 1668	
RESULT 7			
AF339031			
LOCUS			
DEFINITION			
Danio rerio dystrophin (dmd) mRNA, partial cds, alternatively spliced.			
AF339031			
VERSION			
AF339031.1 GI:13699249			
KEYWORDS			
SOURCE			
ORGANISM			
Danio rerio.			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;			
Cypriniformes; Cyprinidae; Danio.			
REFERENCE			
AUTHORS			
Bolanos-Jimenez, F., Bordaies, A., Behra, M., Strahle, U., Sahel, J. and Rendon, A.			
TITLE			
Dystrophin and Dp71, two products of the DMD gene, show a different pattern of expression during embryonic development in zebrafish			
JOURNAL			
Mech. Dev. 102 (1-2), 239-241 (2001)			
MEDLINE			
21184125			
PUBMED			
11287201			
REFERENCE			
AUTHORS			
Bolanos-Jimenez, F., Rendon, A. and Strahle, U.			
TITLE			
Direct Submission			
JOURNAL			
Submitted (22-JAN-2001) Laboratoire de Physiopathologie Retinienne, EMI 99-18 INSERM-Universite Louis Pasteur, 1, place de l'Hopital, Strasbourg 67091, France			
FEATURES			
source			
1..3521			
/organism="Danio rerio"			
/db_xref="taxon:7955"			
/chromosome="1"			
/map="between z5508 and z5058"			
<1..3521			
/gene="dmd"			
<1..3292			
/gene="dmd"			
/note="alternatively spliced"			
/codon_start=2			
/product="dystrophin"			
/protein_id="AAK38376.1"			
/db_xref="GI:13699250"			
/translation="RFRHWEDSHAKLTARVLTQLNMYKSDSDWLEARKRVEPLIKKA NEKLESWKVSHVEDLKGONADVQLSKDLOQWOTQMNVTNELANKLLTLYADDDTS KVKQMTESMLANAWIKKAGDKADLEAGLRQLQHYLYLDLEKFLNLTAEATANVL QDATFKEGLEENPATVRLLEQWDLQAEIDAHREYHSLDENGHRIVSSLEGDNV VLQKRLDDMGORWHELNCNVMSIRPYLDAGVDQWKLHMSLOELLNLWLKRELEKQ KPVGDDVPTVHQQLLTHKAFRELGAKEPVINGTLDNAKTFLEMPREGLKQRPQGD VSPEERVQNVGRILRKEVEDVTVRWKNLGAASVDWQQLLELALERLMELQADQDLY KLRQAESVKNKSWKPVGELLVDLQNHIDRVKAFQEIAPIQDNVNVHVNQLASTFRPSD IQLSPDNLSDLDLNMWRLLQISIEEHLSQLTTAFKDLGSPQNFHASVESPLERSI SPNNVYYINHQTQTTCDWHPKMAELYQLADLNNVRFSAVTAMKLRMMQKALCLDL LSPAAACEAFEQHLNKKQNEQFMDIVQINCLTSYDRLEQHSLSLVNPLCVDMLNW LLNVYDTGRAGKIRTLSEKTIISLCAKLEHLYRFLREVASATGFCQDRLGLLLH DAIQIPQLGEVASFGGSNIEPSVRSQFQFANPKPELEASVFLDWMRLPQSMVWLVP LHRVAAAEATAKHAQCNCKECPITGFYRSLKHFNYDQCQFFSGRVAKGHKMQYP MVEYCTPTSGEDVDFKVLKPKRTKRYFAKHPRMGYLPVQITLEGDNMETPTVLI NFWPVDPHPASSPOLSHDDTHSRIEHYASRDEHLLIQHYCQSLNQSPSLSPQSPAQ ILISMETEKEGELERVLNLEQENRKLQAEYDRLLKKAHDHKLGLSLPLSPQMLPVSPQ SPRDAELIAEAKLLRQHKGRLEARMQILEDHNKQLESQTLRLRLLEQTESKVNQTAL SSPSTASPRSDTSLASLRVAASQTTEIMGDELSSPTQDASTGLEVDIEQLNNSFPHS QGGGRLLNP"			
3'UTR			
3293..3521			
/gene="dmd"			
BASE COUNT			
ORIGIN			
Query Match 46.9%; Score 619.2; DB 5; Length 3521;			
Best Local Similarity 67.9%; Pred. No. 3.6e-169;			
Matches 896; Conservative 0; Mismatches 418; Indels 6; Gaps 2;			
QY	1	CGACTTTCAGCAGTTCAGAAAGCAGATGTACATAGGCGCTTCAAGAGGGAATTGAA	60
Db	844	CGACGTGCCGACCGTTACCAACAACTCCTCAGCAGCAAGGCTTTCAGAAAGAGAGCTGGG	903
QY	61	AACTAAAGAACCTGTATCATGACTCTTGTAGACTGTACGAATATTTCTGACAGAGCA	120
Db	904	TGCCAAAGAACCTGTATCATGACTCTCGACAATGCGAAACCTTCTCTGCTGAGAT	963
QY	121	GCCTTTGGAAGGACTAGAGAAACTCTACCAGGAGCCAGAGAGCTGCCTCCTGAGAGAG	180
Db	964	GCCTCGTGAGGGCCT---GAAACAGAGACCTGGACAAAAGATGTAAGTCTGAGGAGCG	1020
QY	181	AGCCAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGA	240
Db	1021	GTTTCAAAACGTGGGCGCATATTCGCTAAAGAGTGGAGGATGTAACGGTCCGATGGAA	1080
QY	241	AAAAATGAACCTGCACCTCCGCTGACTGCCAGAGAAAAATAGATGAGACCCCTTGAAGACT	300
Db	1081	GAATCTGGGTGCAGCTTCTGTGGACTGGCAGCAACAGCTGGAACCTGGCCTTGGAGAGCT	1140
QY	301	CCAGGAACCTCAAGAGGCCCCACGGATGAGCTGGACCTCAAGCTCGCCAAAGCTGAGGTGAT	360
Db	1141	GATGGAGCTCCAGGACGCCACGAGGATCAGCTGGACTACAAGCTACGACAGGCTGAGAGCGT	1200

QY 361 CAAGGGATCCTGGCAGCCCGTGGGCGATCTCCTCATTTGACTCTCTCCAAAGATCACCTCGA 420
Db 1201 GAAGAATCTTTGGAAGCCTGTTGGGAGCTGCTGTAGACGATTTGCAAAACACATAGA 1260
QY 421 GAAAGTCAAGGCACTTCGAGGAGAAATTCGGCCTCTGAAAGAGAACTGAGCCACGTCAA 480
Db 1261 CAGAGTGAAGGCATTTCAAGAGAGATAGCTCCCATTCAGGATAACGTAAATCATGTTAA 1320
QY 481 TGACCTTGCTCGCCAGCTTACCACCTTTGGGCAATTCAGCTCTCACCCTGATACCTCAGCAC 540
Db 1321 CCAGCTGGCCTCCACATTTAGACCTGCTGACATTCAGCTTCTCCAGACAACTGAGCAG 1380
QY 541 TCTGGAAGACCTGAACACACAGATGGAAGCTTCTGAGGTGGCCGTCGAGGACCGAGTCAG 600
Db 1381 AATCGATGACCTCAACATGAGATGGAGGCTTCTGCAGATCTCCATTCAGGAGCATCTGAG 1440
QY 601 GCAGCTGCATGAAGCCCAACAGGCACTTTGGTCCAGCATCTCAGCACTTTCTTTCCAGTC 660
Db 1441 TCAGCTGACCAACAGCCTTTAAGGACTTGGG---ACCTCTCAGAAATTTCTTTCATCGCTC 1497
QY 661 TGTCCAGGTCCTTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCTACTATATCAACCA 720
Db 1498 TGTGAAAGCCCTCTCGAACGCTCTATTCACCCAAACAAATGTCCTACTATATCAATCA 1557
QY 721 CGAGACTCAAAACAACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCACTTTAGC 780
Db 1558 CCAGACCCAAACAACTGTTGGGACCAACCAAGATGGCAGAACTCTACCACTGATAGC 1617
QY 781 TGACCTGAATAATGTCAGATTTCTCAGCTTATAGGACTGCCATGAACCTCCGAAGACTGCA 840
Db 1618 GGATCTCAACAACTGCGGTTCTCGGCATACAGGACGGCAATGAAGCTCAGACGAATGCA 1677
QY 841 GAAAGCCCTTTGCTTGGATCTCTTGGACCTGTGAGCTGCTGATGATGATGATGATGATG 900
Db 1678 GAAAGCCCTCTGTTGGATCTCTTGGACCTGTGAGCTGCTGATGATGATGATGATGATG 1737
QY 901 CAACCTCAAGCAAAATGACCAAGCCCATGATATCTGCAAGATTAATTAATTTGATGAC 960
Db 1738 CAATCTCAACACAGACAGCAGTTCATGGACATCGTGACAGTGTGATGATGATGATGATG 1797
QY 961 TATTTATGACCGCTGGAGCAAGAGCAACAATTTGGTCAACGTCCTCTCTGCGGTGGA 1020
Db 1798 CATCTACGACCGCTGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1857
QY 1021 TATGTGCTGACTGCTGCTGAATGTTTATGATACGGGACGAAACAGGAGGATCCGTGT 1080
Db 1858 CATGTGCTCAACTGGCTGCTCAACGTTTACGATACAGGACGAGCTGGGAAGATTCGTAC 1917
QY 1081 CCTGTCTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAAGTACAG 1140
Db 1918 CCTATCTTCAAAACAGGAATAATCTTGTGCAAGCTCACTTTGAAGATAAGTACAG 1977
QY 1141 ATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAGCGCAGGCTGGCCT 1200
Db 1978 ATTTTATTTTCGAGAGGTGGCCAGTGGCAGAGCTTCTGTGACCAGCGCCTCGCCT 2037
QY 1201 CCTTCTGCATGATCTATCCAAATTCAGAGCAGTGGGTGAAGTTGCATCCTTTGGGGG 1260
Db 2038 CCTCTGCATGATGCCATTCAGATCCCAGGAGCTGGGTGAAGTGGCGTCTTTCGGAGG 2097
QY 1261 CAGTAACATTGAGCCCAAGTGGCGGAGCTGCTTCCAAATTTGCTAATAAAGCCAGAGAT 1320
Db 2098 GAGCAATATTGAGCCCAAGTGGCGGAGCTGCTTTCAGTTCCGCAATAACAAACCGGAGTT 2157

RESULT 8
HUMMDXX
LOCUS HUMMDXX 2110 bp mRNA linear PRI 07-NOV-1994
DEFINITION Human Duchenne muscular dystrophy (DMD) mRNA, complete cds.
ACCESSION M92650
VERSION M92650.1 GI:181598
KEYWORDS Duchenne muscular dystrophy protein.

SOURCE Homo sapiens brain cDNA to mRNA.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 2110)
AUTHORS Lederfein,D., Levy,Z., Augier,N., Mornet,D., Morris,G., Fuchs,O., Yaffe,D. and Nudel,U.
TITLE A 71-kilodalton protein is a major product of the Duchenne muscular dystrophy gene in brain and other nonmuscle tissues
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (12), 5346-5350 (1992)
MEDLINE 92302239
PUBMED 1319059
FEATURES
source Location/Qualifiers
1..2110
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="Xp21.3-p21.1"
/cell_type="amniotic fluid"
/tissue_type="brain"
1..2110
/gene="DMD"
1..52
/gene="DMD"
/note="G00-119-850"
53..1921
/gene="DMD"
/codon_start=1
/protein_id="AAA52316.1"
/db_xref="GI:181599"
/db_xref="GDB:G00-119-850"
/translation="MREQLKGHEQTQTCWHPKMTLYQSLADLNNVRFSAVRTAMKL
RRLQKALCLDLISAAACDALDQHLNKONDQMDILOINCLTIYDRLEQEHNNLVN
VPLCVDMCLNLLNVYDTGRIRVLSFKTGIISLKAHLEDKYRYLFKQVASSTGF
CDQRLGLLHDSIQIPRLGEVAFSGGNIPEVSRSCFQFANKPEIEAALFLDWMR
LEPQSMVLPVLRVAAETAHQAKCNICKEPLIGFRYSLKHFNYDICQSCFFSG
RVAKGHMHPMVEYCTPTTSGEDVRDEFAVLKNKFKRYFAKHPRMGYLPVQTVLE
GDNMETPASSPQLSHDDTHSRIEYASRLAEMENSGSYLNDISPNESIDDEHLLIQ
HYCQSLNQDSPSPQSPAILISLESEERGERELADLEENRNLAQYDRLKQOH
EHKGLSPSPPEMPTSPQSPRDAELIAEAKLLRQHKGRLEARMQILEDHNKQLESQ
LHRLQLLEQPAEAKVNGTTSVSPSTSLQSDSSQPMLLRVVGSQTSDSMGEDLLS
PPQDSTGLEEVMEQLNNSFPSSRGNVGSFLFHMADDLGRAMESLVSVMTDEEGAE"
1922..2110
/gene="DMD"
3'UTR
BASE COUNT 602 a 511 c 492 g 505 t
ORIGIN
Query Match 45.7%; Score 604; DB 9; Length 2110;
Best Local Similarity 100.0%; Pred. No. 9.2e-165;
Matches 604; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 718 CCACGAGACTCAAACTTGTGGGACCATCCCAAAATGACAGAGCTCTACCACTCTT 777
Db 73 CCACGAGACTCAAACTTGTGGGACCATCCCAAAATGACAGAGCTCTACCACTCTT 132
QY 778 AGCTGACCTGAATAATGTCAGATTTCTCAGCTTATAGGACTGCCATGAACTCCGAAGACT 837
Db 133 AGCTGACCTGAATAATGTCAGATTTCTCAGCTTATAGGACTGCCATGAACTCCGAAGACT 192
QY 838 GCAGAAGGCCCTTTGCTTGGATCTCTTGGCCCTGTGAGCCTGTGATGATGATGATGATG 897
Db 193 GCAGAAGGCCCTTTGCTTGGATCTCTTGGCCCTGTGAGCCTGTGATGATGATGATGATG 252
QY 898 GCACAACCTCAAGCAAAATGACCAAGCCCATGGATATCCTGCAGATTAATTAATTTGAC 957
Db 253 GCACAACCTCAAGCAAAATGACCAAGCCCATGGATATCCTGCAGATTAATTAATTTGAC 312
QY 958 CACTATTTATGACCGCTGGAGCAAGAGCAACAATTTGTCACAGTCCCTCTCTCGGT 1017
Db 313 CACTATTTATGACCGCTGGAGCAAGAGCAACAATTTGTCACAGTCCCTCTCTCGGT 372
QY 1018 GGATATGTCTGACTGGCTGCTGAATGTTTATGATAGGGGACGAACAGGAGGATCCG 1077
|||||

Db 373 GGATATGTCTGAACCTGGCTGCTGAATGTTATGATACGGGACGAACAGGAGGATCCG 432

QY 1078 TGTCTCTGCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTTGGAGACAAAGTA 1137

Db 433 TGTCTCTGCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTTGGAGACAAAGTA 492

QY 1138 CAGATACCTTTTCAAGCAAGTGGCAAGTTCACAGGATTTTGTGACCAGCGCGGCTGGG 1197

Db 493 CAGATACCTTTTCAAGCAAGTGGCAAGTTCACAGGATTTTGTGACCAGCGCGGCTGGG 552

QY 1198 CCTCCTTCTGCATGATTTCTATCCAAATTCACAGACAGTTGGGTGAAGTTGCATCCTTTGG 1257

Db 553 CCTCCTTCTGCATGATTTCTATCCAAATTCACAGACAGTTGGGTGAAGTTGCATCCTTTGG 612

QY 1258 GGGCAGTAACATTTGAGCCAAAGTGTCCGGAGCTGCTTCCAAATTTGCTAATAATAAGCCAGA 1317

Db 613 GGGCAGTAACATTTGAGCCAAAGTGTCCGGAGCTGCTTCCAAATTTGCTAATAATAAGCCAGA 672

QY 1318 GATC 1321

Db 673 GATC 676

RESULT 9

BC028720

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REMARK

COMMENT

Email: cgabps-r@mail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Institute for Systems Biology

http://www.systemsbio.org

contact: amadan@systemsbiology.org

Anup Madan, Jessica Fahey, Erin Helton, Mark Ketterman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAK Plate: 46 Row: j Column: 3

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 5032296.

Location/Qualifiers

1. .4658

/organism="Homo sapiens"

/db_xref="LocusID:1756"

/db_xref="taxon:9606"

/clone="MGC:33083 IMAGE:4822807"

/tissue_type="Testis"

/clone_lib="NIH_MGC_97"

/lab_host="DH10B"

/note="Vector: pBluescript"

130. .2037

/codon_start=1

/product="dystrophin (muscular dystrophy, Duchenne and

CDS

Becker types)"

/protein_id="AAH28720.1"

/db_xref="GI:20379676"

/translation="MREQLKGHETQTTCDWHPKMTLYQSLADLNNVRFSAVRTAMKLRLQKALCLDLLSLSAACDALDQHNKQNDQMDILQIINCLTIIYDRLEQEHNNLVNVPCLVDMCLNWLNVYDTRGRIRVLVSKTGIISLCKAHLEDKYRVLFKQVASSTGFCQRRRLGILLHDSIQIPRLQGEVASFSGSNIEPSCFQFANNKPEIEAALFLDWMRLPQSMVWLPVLHRAAAETAKHQAQCNICKCEPIIGFRYRSLKHFNVDICQSCFFSGRVAKGHKMHYPMVEYCTPTTSGEDVRDFAKVLKNKFKRYFAKHPRMGYLPVQTVLEGNMETPVTLLINFWPVDSAPASSQLSHDDTHSRIEHYASRLAEMENSGSYLNDSSISPNESIDDEHLLIOHYCQSLNODSPLSPSPPEMPTSPQSPRAELIAEAKLLROKHGRLEARNILQAEYDRLKQOHEHKLGLSPLSPPEMPTSPQSPRAELIAEAKLLROKHGRLEARNQILEDHNKQLESQHLRLQLLEQPOAEKAVNGTTVSSPSTSLQSDSSQPMLLRVVGSQTSDSMGEEDLLSPPDSTGTGLEEVMEQLNNSFPSSRGHNVGSLFHMADDLGRAMESLVSMVTDEEGAE"

BASE COUNT 1406 a 955 c 934 g 1363 t

ORIGIN

Query Match 45.7%; Score 604; DB 9; Length 4658;

Best Local Similarity 100.0%; Pred. No. 1e-164;

Matches 604; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 718 CCACGAGACTCAAAACAACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAAGTCTTT 777

Db 150 CCACGAGACTCAAAACAACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAAGTCTTT 209

QY 778 AGCTGACCTGAATAATGTGAGATCTCTTGGATCTCTTGGATCTCTTGGATCTCTTGGATCT 837

Db 210 AGCTGACCTGAATAATGTGAGATCTCTTGGATCTCTTGGATCTCTTGGATCTCTTGGATCT 269

QY 838 GCAGAAAGGCGCTTTGCTTGGATCTCTTGGATCTCTTGGATCTCTTGGATCTCTTGGATCT 897

Db 270 GCAGAAAGGCGCTTTGCTTGGATCTCTTGGATCTCTTGGATCTCTTGGATCTCTTGGATCT 329

QY 898 GCACAAACCTCAAGCAAAATGACAGAGCTCTTGGATCTCTTGGATCTCTTGGATCTCTTGGATCT 957

Db 330 GCACAAACCTCAAGCAAAATGACAGAGCTCTTGGATCTCTTGGATCTCTTGGATCTCTTGGATCT 389

QY 958 CACTATTTATGACCGCTGGAGCAAG 1017

Db 390 CACTATTTATGACCGCTGGAGCAAG 449

QY 1018 GGATATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGAGGATCCG 1077

Db 450 GGATATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGAGGATCCG 509

QY 1078 TGTCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTTGGAGACAAAGTA 1137

Db 510 TGTCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTTGGAGACAAAGTA 569

QY 1138 CAGATACCTTTTCAAGCAAGTGGCAAGTTCACAGGATTTTGTGACCAGCGCGGCTGGG 1197

Db 570 CAGATACCTTTTCAAGCAAGTGGCAAGTTCACAGGATTTTGTGACCAGCGCGGCTGGG 629

QY 1198 CCTCCTTCTGCATGATTTCTATCCAAATTCACAGACAGTTGGGTGAAGTTGCATCCTTTGG 1257

Db 630 CCTCCTTCTGCATGATTTCTATCCAAATTCACAGACAGTTGGGTGAAGTTGCATCCTTTGG 689

QY 1258 GGGCAGTAACATTTGAGCCAAAGTGTCCGGAGCTGCTTCCAAATTTGCTAATAATAAGCCAGA 1317

Db 690 GGGCAGTAACATTTGAGCCAAAGTGTCCGGAGCTGCTTCCAAATTTGCTAATAATAAGCCAGA 749

QY 1318 GATC 1321

Db 750 GATC 753

RESULT 10

MMGUTRPH

LOCUS

DEFINITION

ACCESSION

VERSION

MMGUTRPH

M.musculus mRNA for G-utrophin.

X83506

X83506.1

GI:793851

mRNA

3161 bp

linear

ROD 23-AUG-1995

KEYWORDS G-utrophin.
SOURCE Mus musculus.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 3161)
AUTHORS Blake,D.J., Schofield,J.N., Zuellig,R.A., Gorecki,D.C., Phelps,S.R., Barnard,E.A., Edwards,Y.H. and Davies,K.E.
TITLE G-utrophin, the autosomal homologue of dystrophin Dp116, is expressed in sensory ganglia and brain
JOURNAL proc. Natl. Acad. Sci. U.S.A. 92 (9), 3697-3701 (1995)
MEDLINE 95249540
PUBMED 7731967
REMARK (sites)
REFERENCE 2 (bases 1 to 3161)
AUTHORS Blake,D.J.
TITLE Direct Submission
JOURNAL Submitted (15-DEC-1994) D.J. Blake, Institute of Molecular Medicine, Molecular Genetics, John Radcliffe Hospital, Oxford, Oxfordshire OX3 9DU, UK
FEATURES
source
1..3161
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="AU#3"
/tissue_type="brain"
/clone_lib="adult mouse brain"
/dev_stage="adult"
1..197
198..3161
/function="Unknown"
/standard_name="G-utrophin"
/citation=[1]
/codon_start=1
/evidence=experimental
/product="G-utrophin (predicted protein)"
/protein_id="CAA58496.1"
/db_xref="GI:793852"
/db_xref="MGD:104631"
/db_xref="SPTREMBL:Q61636"
/translation="MQILRLQKCGKLMMAVVRTSLQKVVLHRLQRMVSSPRYO
KLCKDIAEIDAHDNDFKSIDGNQKMKVKGALGNSEATMLQHRLLDMNQWRNDLKA
ASIRAHLEASAEKWNRLASLEELIKWLNKDEELKKQPIGGDVPALQLQYDCHKVL
RELKEKEYSVLNAVQARVPLADQPIEAPEEPRRNQSKTELTPEERAQKIAKMRK
QSSEVREKWNENAVTSNWKQVGVKALEKRLDLQGMDSQSLPLDLHPSLKMSRQD
DLNLSLDHIEKTAFREIAPINLKVKTMDLSSQLPLDLHPSLKMSRQD
RWKLLQVSVDDRLLKQLEAHRDFGSSQHLSTSVQLPWRISHNKVPYIINHQTQ
TQWDHPKMTFLQSLADLNVRFSAYRTAIIKIRRLQKALCLDLLELNTTNEVFKQKL
NONDQLLSVPDVINCLTYYDGLQHLKDLVNPVLCVDMCLNLLNVDYDRTGKIRV
QSLKIGLSLSEKLEEKYRCLFKEVAGPTMCDQRLGLLHLDAIQIPRLGEVAAF
GGSNIEPSVRSFQNNKPEISVKEFIDWMLHLEPQSMVWLPVLRHVAEATAKHQAK
CNICEKPIVGFYRSLKHFNDVQCSCFFSGRTAKGHLHPMVEYCIPTTSGEDVR
DFTKVLKNKFRSKYFAKHPRLGLYPQTVLEGNLETPITLISMWPEHYDPSQSPQL
FHDDTHSRIEQYATRLAQMERTNGSFLTDSSTTGSVEDEHALIQYCYQTLGESPVS
QPQSPAQILKSVEEREGERELRIIADLEEQRNLQVEYEQLEQHLRRLGLPVGSPDPS
IVSPHHTSEDSLEIAEAKLLRQHKRLEARMQILEDHNKQLESQHLRLQLLEQPDSD
SRINGVSPWASPQHSALSYSLDTPGPPQFHQAASEDLLAPPHDTSTDLTDVMEQINST
FPSCSSNVPSRPQAM"

BASE COUNT 887 a 784 c 802 g 688 t

Query Match 43.7%; Score 577.8; DB 10; Length 3161;
Best Local Similarity 65.3%; Pred. No. 4.3e-157;
Matches 868; Conservative 0; Mismatches 452; Indels 9; Gaps 1;

QY 2 GACTTCCAGAGTTCAGAGCAGACGATGATACATAGGCGCTTCAAGAGGGAATTGAAA 61
Db 633 GACGTCCCTGCTTACAGCTCCAGTATGACCACTGCAAGGTGCTGAGACGTGAGCTAAAG 692
QY 62 ACTAAGAACCTGTAAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGACGAG 121
Db 693 GAGAAAGAGTATCTGTGCTGAACGCCCTAGATCAAGCTCGAGTTTTTCTGCTGATCAG 752

QY 122 CCTTT-----GGAAGGACTAGAGAAAACCTCTACCAGGAGCCCGAGAGAGCTGCCTCCT 172
Db 753 CCAATAGAGGCCCCGGAAGAACCAAGAAACCCACAATCAAGAGCAGAGTTGACTCCT 812
QY 173 GAGGAGAGAGCCCGAGAAATGTCACCTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACT 232
Db 813 GAGGAGAGAGCCCGAGAAATGTCACCTCGGCTTCTACGAAAGCAGGCTTCTGAGTCCGAGAG 872
QY 233 GAGTGGGAAAAATTGAACCTGCACTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTT 292
Db 873 AAGTGGGAAAAATCTAAATGCTCACTAGCAACTGGCAAAAGCAAGTAGGGAAGGCGTTA 932
QY 293 GAAAGACTCCAGGAACCTCAAGAGAGCCACGGATGAGCTGGACCTCAAGCTGGCCAAAGCT 352
Db 933 GAGAAACTCCGAGACCTTGCAGGGAGCTATGGACGACCTGGACGACATGAAGGAGGTG 992
QY 353 GAGGTGATCAAGGGATCTTGGCAGCCCGTGGCGGATCTCCTCATTTGACTCTCTCCAAGAT 412
Db 993 GAGGCTGTGCGGAATGGCTGGAAGCCCGTGGGAGACCTGCTTATAGACTCCCTGCAGGAT 1052
QY 413 CACCTCGAGAAAAATCAAGGCACCTTCGAGGAGAAATTCGCGCTCTGAAAGAGAACGCTGAGC 472
Db 1053 CACATCGAGAAAAACCTTGGCGTTTAGAGAGAAATTCACCAATCAACTTAAAAAGTAAAA 1112
QY 473 CACGCTCAATGACCTTGTCTGCCAGCTTACCACCTTTGGGCATTCAGCTCTCACCGTATAAC 532
Db 1113 ACAATGAATGACCTGTCCAGTCCAGTGTCTCCACTTGACTTGCATCCATCTCTAAAGATG 1172
QY 533 CTCAGCACTCTGGAAGACCTGACACCAACAGATGGAAGCTTTCGAGGTGGCGCTCGAGGAC 592
Db 1173 TCTCGCCAGCTGGATGACCTTAATATGCGATGGAACCTTCTACAGGTTCCGTGGACGAT 1232
QY 593 CGAGTCAGGCAGCTGCATGAAGCCACAGGACTTTGGTCCAGCATCTCAGCACTTTTCTT 652
Db 1233 CGCCTTAAGCAGCTCCAGAAAGCCACAGAGATTTTGGGCCATCTTCTCAACACTTTCTG 1292
QY 653 TCCACGCTGTCTCCAGGCTCCCTGGGAGAGAGCCATCTCGCAAAACAAAGTGCCCTACTAT 712
Db 1293 TCCACTTCAGTCCAGCTGCCGTGGCAGAGATCCATTTCACATAATAAAGTGCCCTATTAC 1352
QY 713 ATCAACCCAGGAGACTCAAAACAACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAG 772
Db 1353 ATCAACCATCAAAACACAGACAACTGTTGGGATCATCTTAAATGACTGAGCTCTTCCAA 1412
QY 773 TCTTAGCTGACCTGAATAATGTACAGATTCAGATTCAGCTTATAGGACTGCCATGAAACTCCGA 832
Db 1413 TCCCTTGTGATCTGAATAATGTACGTTTCTCTGCTTCTGCTGAGCTGATGATGATGATG 892
QY 833 AGACTGCAGAAAGGCCCTTTGCTTGGATCTCTTGGAGCTGTGAGCTGATGATGATGATGATG 952
Db 1473 AGGCTGCAAAAGCATTATGTCTGGATCTCTTAGAGCTGAATACGACGAAATGAAGTTTC 1532
QY 893 GACGACACAACTCAAGCAAAATGACAGCCCATGGATATCTTGCAGATATCTTGAATATTAATGT 952
Db 1533 AAGCAGCACAACCTGAACCAAAATGATCAGCTCCTGAGTGTCCCAGACGCTCATCAACTGT 1592
QY 953 TTGACCACTATTTATGACCGCTGGAGCAAGAGCACAACAATTTGGTCAACGCTCCCTCTC 1012
Db 1593 CTGACCACTTACGATGGCTTGGAGCTGACAGCTGCACAAGGACCTTGGTCAATGTTCCACTC 1652
QY 1013 TCGGTGGATATGTGTGAACTGGCTGCTGAATGTTTATGATACGGGACGACAGGAGG 1072
Db 1653 TCGGTGATATGTGTCAACTGGCTGCTCAACGCTATACGACACGCGGCGGACTGGAATA 1712
QY 1073 ATCCGTGTCTCTTTTAAACTGGCATCTTCCCTCTGTAAGACACATTTTGAAGAGAC 1132
Db 1713 ATTCGGGTACAGATCTGAAGATTGGATTGATGCTCTCTCTCCAAAGGCTCTTTAGAGAG 1772
QY 1133 AAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCGAGGAG 1192
Db 1773 AATAACAGATGTCTCTTTAAGGAGGTGGCAGGCGCAACTGAGATGTGTGACCGAGGAG 1832
QY 1193 CTGGGCTCCTTCTGTCATGATCTATCCAAATTCGAAGACAGTTGGGTGAAGTTGCATCC 1252


```
QY 1034 TGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCGTCCTGTCTTTTAAA 1093
Db 1750 TGGCTCCTCAATGTTTATGATAGTGGTCGACGGGAAAGATCGGGCATTTGCTTTAAG 1809
QY 1094 ACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAAGTACAGATACCTTTTCAAG 1153
Db 1810 ACTGGCATTCATGCTTGTGTGGCACGGAAGTGAAGGAAACCTTCAGTACCTCTTCAGC 1869
QY 1154 CAAGTGGCAAGTCAACAGGATTTTGTGACGAGCGCAGGCTGGGCCCTCTCTTCATGAT 1213
Db 1870 CAAGTGGCAAGTCAACAGGATTTTGTGACGAGCGCAGGCTGGGCCCTCTCTTCATGAT 1213
QY 1214 TCTATCCAAATTCGAAGACAGTGGGTGAAGTTCATCCTTTGGGGGCGAGTAAATGAG 1273
Db 1930 GCCATTCAGGTGCCCCGTCAGCTGGGTGAAGTGGCAGCCTTTGGGGGCGAGTAAATGAG 1273
QY 1274 CCAAGTGTCCGAGGCTGCTTCCAAATTTGCTAATAATAAGCCAGAGAT 1320
Db 1990 CCCAGTGTCCGAGTGTGCTCCGTTTACGACCGGGGAGCCAGTCAAT 2036

RESULT 12
AF195788
LOCUS
DEFINITION
Rattus norvegicus 3073 bp mRNA linear ROD 02-MAY-2001
variant (Drp2) mRNA, complete cds.
AF195788
AF195788.1 GI:11066166
SOURCE
Rattus norvegicus.
ORGANISM
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 3073)
Roberts, R.G. and Sheng, M.
Association of dystrophin-related protein 2 (DRP2) with
postsynaptic densities in rat brain
Mol. Cell. Neurosci. 16 (5), 674-685 (2000)
20538711
PUBMED
11083927
REFERENCE
2 (bases 1 to 3073)
Roberts, R.G. and Sheng, M.
Direct Submission
Submitted (18-OCT-1999) Division of Medical and Molecular Genetics,
GKT Medical School, 8th Floor, Guy's Hospital, Guy's Tower, London
SE1 9RT, UK
FEATURES
source
1. .3073
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
1. .3073
/gene="Drp2"
69. .70
/misc_feature
/gene="Drp2"
/note="Drp2"
/note="splice acceptor AG dinucleotide used by A-form"
200. .3073
CDS
/note="Drp2"
/note="DRP2; membrane-associated cytoskeletal protein;
contains two spectrin repeats; WW domain; Z2 domain"
/codon_start=1
/product="dystrophin-related protein 2 B-form splice
variant"
/protein_id="AAG28485.1"
/db_xref="GI:11066167"
/translation="MQPLVMQGPVTLPRCHEWHAADRHHSSSLRNTCPQPQVRAAV
TIPAPWDGADPCLSPKLLNGSVGAVGPLEPSAMNLCWNEIKKSHNLRLEAFSD
HSGKLQPLQEIIDWLSQKDELSAQLPLOGDVALVQKEKETHAAFMEEVKSGPYIY
SVLESQAFLSQHPFEELESSESKDTSRQRIQNLRSFVWKQATVASELWEKLTAR
CVDQHRHIEHTLEHLLIQAMEELSSLTQAEGVRAIWEPIGDLFIDSLPEHIQAIK
LFKEEFSVPKDGKLVNLDLAHQLAISDVHLSMENSRALEQINVRWKQLQVSAERLKO
```

```
LODAHRDFGPGSQHFLSTSVQVPWERAISPKNKPYIINHQAQTTQWDHPKMTELYQTL
ADLNKIFSAVRTAMKLRVQKALRLDLVTLTALTEIFNEHDLQASEHVMVVEVIHC
LTALYERLEERGIILNVPLCVDMSLNLLNVDGSRSGKMRALSFKTGLIACLCGTEV
KEKLYLFQSVANSGSKDQRLHALLHEATQVPRQLGEVAAFGGSNVESVRSFCFRF
STGKPVIEASQFLEWVNLPEQSMVWLAVLHRVTVAEQVHKQTKSCICROCPKIGFRYR
SLKQFNVDICQTCFLTGRASKGNKLHYIMEYTYPTTSSENMRDEATLKNKFRSKQY
FSKHPRGYLPVQSVLEDCSETPASSPMLPHADTHSRIFHEFASRLAEESQNCSEFFN
DSLSPDDSIDEDQYLLRHSSPIDREPAFGQQAFCMSMATESKGELEKILAHLEDENRI
LOGELRLKWOHEEAVEAPTLAGSAEATPDHRNEELAEARILRQHKSRLETRMQIL
EDHNQLESQQLRLRELLQPTESDNGSAGSSLASSESGSHPREKQTTPTDTE
AADDVGSKSQDVSHSLEDIMEKLRHAFPSVRSDDVTANTLLAS"
BASE COUNT 779 a 843 c 790 g 659 t 2 others
ORIGIN
```

```
Query Match 41.8%; Score 551.8; DB 10; Length 3073;
Best Local Similarity 63.9%; Pred. No. 1.7e-149;
Matches 835; Conservative 0; Mismatches 472; Indels 0; Gaps 0;
```

```
QY 14 GTTCAGAAAGCAGAACGATGTACATAGGCGCTTCAAGAGGGAATGAAAACTAAAGAACCT 73
Db 611 GTACAAACAGGAGAGGAGACACATGCTGCCTTTATGGAAGAGTGAAGTCTAAAGGCCCC 670
QY 74 GTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAGGA 133
Db 671 TATATCTACTCTGTGCTGGAATCGGCACAGGCTTCTCTGTCAGCACCCCATTTGAGGAA 730
QY 134 CTAGAGAAACTCTACCAGGAGCCCGAGAGAGTGCCTCTGAGGAGAGAGCCGAGAAATGTC 193
Db 731 TTAGAGGAGTCTCATTTCTGAGAGCAAGATACCTCCCCAAGACAGCGGATTCAAAACCTC 790
QY 194 ACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATCTAGTGGGAAAAATTTGAACCTG 253
Db 791 AGTCGCTTGTGGAAGCAGGACAGTGGCCAGTGAACCTGAGGAGAGTTCAAAACCTC 790
QY 254 CACTCCGCTGACTGGCAGAGAAAATAGATGAGACCCCTTGAAGACTCCAGGAACTTCAA 313
Db 851 CGCTGTGTGATCAGCATCGCACATGAGCACACCCCTGGAACATCTATTGGAGATCCAA 910
QY 314 GAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCAGCTGAGGTGATCAAGGGATCCTGG 373
Db 911 GGGGCAATGGAGAACTGAGCAGTACTTTGACCCAGCAGAGGGAGTCCGAGCCACATGG 970
QY 374 CAGCCCGTGGCGCATCTCTCATTTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGCA 433
Db 971 GAGCCCATAGGAGATCTTTATCGATTCTCTCCAGAGCATATCCAAGCATCAAGTTA 1030
QY 434 CTTCGAGGAGAAATTCGCCCTCTGAAAGAGAACGTGAGCCACGTCAATGACCTTGTCTGC 493
Db 1031 TTCAAAGAAGAAATCTCTCTGTGAAAGACGGGTGAAGTTAGTGAATGACCTGGCCAC 1090
QY 494 CAGCTTACCACCTTTGGGCATTCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCTG 553
Db 1091 CAGCTTGTCTATTCTGTGATGTGCATCTGCAATGGAGATTCGAAGGGCTCTGGAACAGATC 1150
QY 554 AACACCAGATGGAAGCTTCTGAGGTGGCGGTGAGGACCGAGTCAAGCAGCTGCATGAA 613
Db 1151 AACGTCCGCTGGAACAGCTCCAGGTGTGAGTGTGCTGAGAGGCTTAAACAGCTCCAGGAT 1210
QY 614 GCCCAGAGGGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGTCTGTCCAGGGTCCC 673
Db 1211 GCCCAGCGGACTTTGGGCTTGGGTTCACAGCACTTCTCTCCACTTCTGTTCAAGTTCCC 1270
QY 674 TGGGAGAGAGCCATCTCGCCAAACAAAGTGCCTTACTATATCAACCAGAGACTCAAACA 733
Db 1271 TGGGAAAGAGCAATTTCTCCCAATAAAGTCCCTACTACATCAACACCAGGCTCAGACC 1330
QY 734 ACTTGTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAAT 793
Db 1331 ACATGCTGGGACCATCCCAAGATGACCGAGTTATACCAACCCCTAGCTGACCTGAACAAC 1390
QY 794 GTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAGAGCTGCAGAGGCCCTTTCG 853
Db 1391 ATCAAGTTCTCAGCTTACCGCACTGCCATGAAGTCCCGGAGTCCAGAGGCCCTGCGC 1450
```


Mon Apr 28 09:28:22 2003

variant"
/protein_id="AAG28484.1"
/db_xref="GI:11066165"
/translation="MQPLVMQGCPTLPRCHEWHAADRFHSSSLRNTCPQPVRAAV
/TIPAPWDGAGDPCLSPKLLNGSVGAVGLEPSEAMNLCWNEIKKSHNLARLEAFSD
HSGKIQPLQEIIDLWLSQKDELSAQLPQGDVALVQOEKETHAFAFMEEVKSKGPIYI
SVLESAQFLSQHPFEELESHSEKDTSPQRIONLSRFVWKQATVASELWEKLTAR
CVDQHRHIEHTLEHLEIQGMEELSSTLTQAEVGRATWEPIDGFLTDSLPEHIOAIK
LFKEEFSVKGDKVNLDAHQALASDVHLSMENSRALEQINVRWKQLQVSVARLKO
LQDAHRDQPGSGHFLSTSVQVFWERAI SPNKVPYVINHQATQTCWDHPKMTLEYQTL
ADLNNIKFSAYRTAMKLRVQKALRLDLVTLTALEIFNEHDLQASEHVMVDVEVHC
LTALEYERLEEEERGLVNPVLCVDMSLNWLNVFDSGRSGKMRALSFKTGACLCGTEV
KEKLYLFSSQVANSKCDQRHLGALLHEAIQVPRQLGEVAAFGGSNVEPSVRSFRR
STGKPVIEASQFLEWNLPEPQSMVWLAVLHRTVAVQVKHQTCSICRQCPKGRYR
SLKQFNVDICQTCFLTGKASKGNKLHYPIMEYITPTSSNMRFATTLKNKFRSKOY
FSKHPORGYLPVQSVLESDCSETPASSPMLPHADTHSRIEHFASRLAEMESQNCFFN
DSLSPDDSIDEDQYLLRHSSPITDRPAFGQQAQPCSMATESKGELEKILAHLEDENRI
LQGLRLRLKWQHEEAVEAPTAEQSAEATPDHRNEELAEARILRQHKSRLETRMOIL
EDHNKQLESQLORLRELLLOPPTESDNGSAGSSSLASSPROSEGSHPREKGQTPDTE
AADDVGSKSQDVSHSLEDIMEKLRHAFPSVRSDDVTANTLLAS"
BASE COUNT 790 a 864 c 794 g 690 t 2 others
ORIGIN
Query Match 41.8%; Score 551.8; DB 10; Length 3140;
Best Local Similarity 63.9%; Pred. No. 1.7e-149;
Matches 835; Conservative 0; Mismatches 472; Indels 0; Gaps 0;
QY 14 GTTCAGAAAGCAGAACGATGTACATAGGGCCCTTCAAGAGGGAATTTGAAACCTAAAGAACCT 73
Db 678 GTACAACAGGAGAGGAGACACATGCTGCTTTATGGAAGAAGTGAAGTCTAAAGGCCCC 737
QY 74 GTAATCATGAGTACTTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGAAGGA 133
Db 738 TATATCTACTCTGCTGGAATCGGCACAGGCTTCTCTGTCTCAGCACCCATTTGAGGAA 797
QY 134 CTAGAGAACTCTACCAGGAGCCCGAGAGCTGCTCTCTGAGGAGAGAGCCCGAATGTC 193
Db 798 TTAGAGGAGTCTCATTTCTGAGAGCAAGATACCTCCCAAGACAGCGGATTCAAAACCTC 857
QY 194 ACTCGGCTTTACGAAAGCAGGCTGAGGAGGTCAATCTGAGTGGGAAAAATTTGAACCTG 253
Db 858 AGTCGCTTTGTGTGGAAGCAGGCAACAGTGGCCAGTGAACCTGAGGAGAGCTGACAGCC 917
QY 254 CACTCCGCTGACTGGCAGAGAAAATATAGATGAGACCCCTTGAAGACTCCAGGAACCTCAA 313
Db 918 CGCTGTGTGATCAGCATCGCCACATTTGAGCACACCCCTTGAACATCTATTGGAGATCCA 977
QY 314 GAGGCCACGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGATCCTGG 373
Db 978 GGGGCAATGGAGGAAGTGAAGTGAACCTTGTGACCAAGCAGAGGGAGTCCGAGCCACATGG 1037
QY 434 CTTCGAGGAGAAATTTGGCCCTCTGAAAAGAGAACGTCGACCCAGTCAATGACCTTGCTCGC 493
Db 1098 TTCAAAGAAGAATTTCTCTCTCTGAAAAGACGGGTGAAGTAGTGAATGACCTGGCCAC 1157
QY 494 CAGCTTACCACCTTTGGGCATTTGAGCTCTCAGCTCTCACCGTATAACCTCAGCAGCTCTGGAAGACCTG 553
Db 1158 CAGCTTGTATTTCTGATGTGCACCTTGTCAATGGAGAATTTCAAGGGCTCTGGAACAGATC 1217
QY 554 AACACCAGATGGAAGCTTTCTGAGGTGGCGCTCGAGGACCGAGTCAGGAGCTGCATGAA 613
Db 1218 AACGTCCGCTGGAACACAGCTCCAGGTGTGAGTGTGCTGAGAGGCTTAAACAGCTCCAGGAT 1277
QY 614 GCCACACAGGACTTTGGTCCAGCATCTCAGCAGCTTCTTCCAGCTCTGTCCAGGGTCCC 673
Db 1278 GCCCACCAGGACTTTGGGCTGGGTTCACAGCAGCTTCTCTCCACTTCTGTTCAAGTCCC 1337
QY 674 TGGGAGAGAGCCATCTCGCAACAAAGTGCCCTACTATATCAACACGAGACTCAACA 733

854 TTGGATCTCTTGGCCCTGTGATGTCAGCTGATGTCGCTTGACCCAGCAGCACACCTCAAGCAA 913
1451 TTGGATCTGGTAACCTTAACTACAGCTCTGGAGATCTTCAATGAGCAGCAGCTGCGAGGCC 1510
914 AATGACACAGCCCATGGATATCTCTGCAGATTATTAAATGTTTGGACCACTATTTATGACCCG 973
1511 AGTGAACATGTAATGGATGGTGGAGGTCAATTCATCTGCTGAGCTGCTTGTATGAACGA 1570
974 CTGGAGCAAGAGCACAACAATTTGGTCAACGTCCTCTCTGCGGTGGATATGTGCTGAAC 1033
1571 CTGGAGGAGAAAGAGGCATCCTGGTCAACGTCGCGCTGTGTGATAGACATGAGCCTCAAC 1630
1034 TGGCTGTGTAATGTTATGATACGGGAGCAACAGGAGGATCCGCTGCTCTTTTAA 1093
1631 TGGCTCCTCAACGTTTGTAGATGGCCGAGTGGAAAGATGCGAGCATTGTCTTTAAG 1690
1094 ACTGGCATCATTTCCCTGTGTAAAGCACATTTTGAAGACAAAGTACAGATACCTTTTCAAG 1153
1691 ACTGGCATCGCATGCTGTGGCACTGAAGTGAAGAAACCTTCAGTATCTCTTTCAGC 1750
1154 CAAGTGGCAAGTTCAACAGGATTTTGTGACACGCGCAGGCTGGGCTCTCTTGCATGAT 1213
1751 CAAGTGGCAATTCAGGCAAGCAAGTGTGACACGCGCCCTCGGTGCGCTTCTCATGAA 1810
1214 TCFATCCAAATTTCCAAAGACAGTGGTGAAGTTGCATCCTTTGGGGGCGAGTAACATGAG 1273
1811 GCCATCCAAGTGCCCGCTGAGTGGTGAAGTGGCAGCATTTGGGGGCGAGCAATGTGGAG 1870
1274 CCAAGTGTCCGAGCTGCTTCCAAATTTGCTAATAATAAGCCAGAGAT 1320
1871 CCCAGTGTCCGTAGTGTGCTTTCGCTTCCAGCTGGCAAGCCAGTGCAT 1917
RESULT 13
AF195787 3140 bp mRNA linear ROD 02-MAY-2001
LOCUS Rattus norvegicus dystrophin-related protein 2 A-form splice
DEFINITION variant (Drp2) mRNA, complete cds.
ACCESSION AF195787.1 GI:11066164
VERSION AF195787
KEYWORDS Rattus norvegicus.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
REFERENCE 1 (bases 1 to 3140)
AUTHORS Roberts, R.G. and Sheng, M.
TITLE Association of dystrophin-related protein 2 (DRP2) with postsynaptic densities in rat brain
JOURNAL Mol. Cell. Neurosci. 16 (5), 674-685 (2000)
MEDLINE 20538711
PUBMED 11083927
REFERENCE 2 (bases 1 to 3140)
AUTHORS Roberts, R.G. and Sheng, M.
TITLE Direct Submission
JOURNAL Submitted (18-OCT-1999) Division of Medical and Molecular Genetics, GKT Medical School, 8th Floor, Guy's Hospital, Guy's Tower, London SE1 9RT, UK
FEATURES
source Location/Qualifiers
1. .3140
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
1. .3140
/gene="Drp2"
267. .3140
/note="Drp2"
/note="Drp2; membrane-associated cytoskeletal protein; contains two spectrin repeats; WW domain; ZZ domain"
/codon_start=1
/product="dystrophin-related protein 2 A-form splice"

Db	1338	TGGGAAAGAGCAATTTCTCCCAATAAAGTCCCTACTACATCAACCACAGGCTCAGACC	1397
QY	734	ACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAAT	793
Db	1398	ACATGCTGGGACCATCCCAAGATGACCGAGTTATACCAAAACCTAGCTGACCTGAACAAC	1457
QY	794	GTGAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAAGGCCCTTGC	853
Db	1458	ATCAAGTTCTCAGCTTACCGCACTGCCATGAAGCTCCGCCGAGTCCAGAAAGGCCCTGCGC	1517
QY	854	TTGGATCTCTTGAGCCTGTGACGCTGCATGTGATGCTTGGACCAAGCAACACCTCAAGCAA	913
Db	1518	TTGGATCTGTTAACTTAACTACAGCTCTGGAGATCTTCAATGAGCAGCAGCTGACGGCC	1577
QY	914	AATGACCAAGCCCATGGATATCCTGCAGATTTATTAATTTGTTGACCACTATTTATGACCGC	973
Db	1578	AGTGAACATTAATGGATGTGGTGGAGGTCATTCACCTGCTTGACTGCTTGTATGAACGA	1637
QY	974	CTGGAGCAAGAGCACAACAATTTGGTCAACGTCCTCTCTGCTGCTGGATATGTCTGAAC	1033
Db	1638	CTGGAGGAGGAAGAGGCATCTGCTCAACGTCGCGCTGTGTGTAGACATGAGCCTCAAC	1697
QY	1034	TGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCGTCCTGCTCTTTTAA	1093
Db	1698	TGCTCTCTCAACGTTTTTGTATAGTGGCCGAGTGGAAGATGCGAGCATTTGCTCTTAAG	1757
QY	1094	ACTGGCATCATTTCCCTGTGTTAAAGCACATTTGGACCAAGCAAGTACAGATACCTTTTCAAG	1153
Db	1758	ACTGGCATCGCATGCTGTGTGGCACTGAAGTGAAGAAAGAAACTTCAGTATCTCTCAGC	1817
QY	1154	CAAGTGGCAAGTTCAACAGGATTTTGTGACCAAGCGCAGGCTGGGCTCTCTGTCATGAT	1213
Db	1818	CAAGTGGCAATTCAGGCAGCAAGTGTGACCAAGCGCCACCTCGGTGCTCTCATGAA	1877
QY	1214	TCTATCCAAATTCGAAGACAGTGGTGAAGTTGCATCCTTTGGGGGAGTAACATTTGAG	1273
Db	1878	GCCATCCAAAGTGCCCGCTCAGCTGGTGAAGTGGCAGCATTTGGGGGAGCAATGTGGAG	1937
QY	1274	CCAAGTGTCCGGAGCTGCTTCCAAATTTGCTAATAATAAGCCAGAGAT	1320
Db	1938	CCAGTGTCCGTAGTTGCTTTCGCTTCAGCACTGGCAAGCCAGTCAT	1984
RESULT 14			
RNAPDY3			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
FEATURES			
source			
Location/Qualifiers			
1..1240			
/organism="Rattus norvegicus"			
/db_xref="taxon:10116"			
/cell_line="Schwannoma cells"			
1240 bp			
R.norvegicus mRNA for APO-dystrophin-3.			
X69767			
X69767.1 GI:57913			
apodystrophin.			
Rattus norvegicus.			
Rattus norvegicus.			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;			
Rattus.			
1 (bases 1 to 1240)			
Tinsley,J.M.			
Direct Submission			
Submitted (14-DEC-1992) J.M. Tinsley, Molecular Genetics Group,			
Institute of Molecular Medicine, John Radcliffe Hospital,			
Headington, Oxford OX3 9DU, UK			
2 (bases 1 to 1240)			
Tinsley,J.M., Blake,D.J. and Davies,K.E.			
Apo-dystrophin-3: a 2.2kb transcript from the DMD locus encoding			
the dystrophin glycoprotein binding site			
Hum. Mol. Genet. 2 (5), 521-524 (1993)			
93299455			
8518789			

CDS		/dev_stage="neonatal"	
		106..1128	
		/codon_start=1	
		/product="apodystrophin-3"	
		/protein_id="CAA49423.1"	
		/db_xref="GI:57914"	
		/db_xref="SPTREMBL:Q05485"	
		/translation="MREHLKGHEQTTCWDPKMTLYQSLADLNNVRFSAYRTAMKL	
		RRLQKALCLDLISLAAACDALDQNLKQNDQMDILQIYYCLTIYDRLEQEHNNLVN	
		VPLCVDVCLNLLNVDYDTRGRIRVLSFKTGIISLKAHLEDKRYLKFQVASSSTGF	
		CDQRLGLLLHDSIQIPROLGEVAFSGSNIIEPVSFSCFQFANNKPEIEAALFLDWMR	
		LEPQSMVWLPVLRVAAAETAHQAQKCNICEKPIIGFRYSRLKHFNYDICQSCFFSG	
		RVAKGHKMHPMVEYCTPTTSGEDVDFAKVLKNKFKTRKRYFAKHPRMGYLPVQTVLE	
		GDNMET"	
BASE COUNT		323 a	287 g
ORIGIN		314 c	316 t
Query Match			
Best Local Similarity 41.2%; Score 544; DB 10; Length 1240;			
Matches 580; Conservative 0; Mismatches 60; Indels 0; Gaps 0;			
QY	681	GAGCCATCTCGCCAAACAAAGTGCCTACTATATCAACCACGAGACTCAACAACTTGT	740
Db	89	GTGAATCCTTACAACCATGAGGGAACACCTCAAGGCCACGAGACCCAACTTGT	148
QY	741	GGGACCATCCCAAAATGACAGACTCTACCACTCTTTAGCTGACCTGAATAATGTCAGAT	800
Db	149	GGGACCATCCCAAAATGACAGACTCTACCACTCTTTAGCTGACCTGAATAATGTCAGAT	208
QY	801	TCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAAGGCCCTTTGCTTGGATC	860
Db	209	TCTCCGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAAGGCCCTTTGCTTGGATC	268
QY	861	TCTTGGCCTGTGACGTGCATGTGATGCCTTGACCAAGCAACCTCAAGCAAAATGACC	920
Db	269	TCTTGGCCTGTGACGTGCATGTGATGCCTTGACCAAGCAACCTCAAGCAAAATGACC	328
QY	921	AGCCCATGGATATCCTGCAGATTTATTAATTTGACCACTATTTATGACCGCTGGAGC	980
Db	329	AGCCCATGGATATCCTGCAGATTTATTAATTTGACCACTATTTATGACCGCTGGAGC	388
QY	981	AAGAGCACAAATTTGGTCAACGTCCTCTCTGCGTGGATATGTCCTGAACCTGGCTGC	1040
Db	389	AAGAGCACAAATTTGGTCAACGTCCTCTCTGCGTGGATATGTCCTGAACCTGGCTGC	448
QY	1041	TGAATGTTTATGATACGGGACGAACAGGAGGATCCGTCCTGTCTTTTAAACTGCCA	1100
Db	449	TCAATGTTTATGATACGGGACGAACAGGAGGATCCGTCCTGTCTTTTAAACTGCCA	508
QY	1101	TCATTCCCTGTGTAAGACACATTTGGAAGACAGTACAGTACATCTTCAAGCAAGTGG	1160
Db	509	TCATTCCCTGTGTAAGACACATTTGGAAGACAGTACAGTACATCTTCAAGCAAGTGG	568
QY	1161	CAAGTTCACAGGATTTGTGACCAAGCGCAGGCTGGGCTCTTCTGCTGATTTCTATCC	1220
Db	569	CAAGTTCACAGGATTTGTGACCAAGCGCAGGCTGGGCTCTTCTGCTGATTTCTATCC	628
QY	1221	AAATTCACAGGATTTGGTGAAGTTGCATCTTTGGGGGAGTAAACATTTGAGCCAAAGT	1280
Db	629	AAATTCACAGGATTTGGTGAAGTTGCATCTTTGGGGGAGTAAACATTTGAGCCGAGT	688
QY	1281	TCCGAGCTGCTTCCAAATTTGCTAATAATAAGCCAGAGAT	1320
Db	689	TCAGGAGCTGTTTCCAGTTTGCCTAATAATAAGCCAGAGAT	728

RESULT 15
E30222
LOCUS
E30222
DEFINITION
Shortened dystrophin.
E30222
ACCESSION
E30222
VERSION
E30222.1 GI:13017029
KEYWORDS
JP 1999318467-A/5.

3172 bp
DNA
linear
PAT 18-JUN-2001

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 07:25:57 ; Search time 184.2 Seconds
(without alignments)
16150.348 Million cell used

Title: US-09-845-416-10_COPY_1800_3120

Perfect score: 1321

Sequence: 1 cgactttccagcagttcaga.....ctaataataagccagagatc 1321

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4313688

Minimum DB seq length: 0

Maximum DB seq length: 5000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

```
Database :
N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
```

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1321	100.0	1821	24	AAD37241	Human dystrophin r
2	1321	100.0	2169	24	AAD37232	Human dystrophin r
3	1321	100.0	3510	24	AAD37240	Human dystrophin m
4	1321	100.0	3531	24	AAD37238	Human dystrophin m
5	1321	100.0	3858	24	AAD37237	Human dystrophin m
6	1321	100.0	3999	24	AAD37234	Human dystrophin m
7	1321	100.0	4182	24	AAD37230	Human dystrophin m
8	1321	100.0	4476	24	AAD37259	Adeno-associated v
9	1321	100.0	4498	24	AAD37258	Adeno-associated v

10	1321	100.0	4825	24	AAD37257	Adeno-associated v
11	1321	100.0	4848	24	AAD37263	Adeno-associated v
12	1321	100.0	4966	24	AAD37256	Adeno-associated v
13	1321	100.0	4990	24	AAD37262	Adeno-associated v
14	1105.2	83.7	4402	21	AAZ48567	A rod shortened dy
15	1037.6	78.5	3446	24	AAD37242	Human dystrophin m
16	1037.6	78.5	4414	24	AAD37260	Adeno-associated v
17	1036	78.4	1434	24	AAD37243	Human dystrophin r
18	906	68.6	4402	21	AAZ48568	Human dystrophin r
19	770.4	58.3	4075	21	AAZ48569	A rod shortened dy
20	766.2	58.0	3747	21	AAZ48566	A rod shortened dy
21	712.6	53.9	3163	21	AAZ48571	A rod shortened dy
22	493.8	37.4	3172	21	AAZ48570	A rod shortened dy
23	492	37.2	887	24	ABK81993	Human dystrophin C
24	324	24.5	324	24	ABK81991	Human dystrophin S
25	289	21.9	387	24	ABK81990	Human dystrophin S
26	215	16.3	215	24	ABK81992	Human dystrophin h
27	160.6	12.2	4928	23	ABL12611	Drosophila melanog
28	114	8.6	114	24	ABK82003	DNA encoding dyster
29	54.2	4.1	2247	24	ABK70286	Human lung cancer
30	44.2	3.3	2574	24	ABQ62254	Mouse syntaxin 4 i
31	42.8	3.2	1690	22	AAI58193	Human polynucleoti
32	41.2	3.1	2105	22	AAS34373	Human cDNA encodin
33	38.6	2.9	1995	12	AAQ14184	N.clavipes draglin
34	38.6	2.9	1995	19	AAV23250	Nephila clavipes s
35	38.6	2.9	1995	21	AAZ38196	N. clavipes spider
36	38.6	2.9	3045	23	ABL14199	Drosophila melanog
37	38.2	2.9	2518	23	AAS73627	DNA encoding novel
38	38.2	2.9	4439	24	ABQ62261	Human syntaxin 4 i
39	37.6	2.8	944	22	AAI88876	Human polynucleoti
40	37.6	2.8	1688	24	ABK64523	Human benign prost
41	37.2	2.8	891	22	AAH67722	C glutamicum codin
42	37.2	2.8	960	22	AAH67723	C glutamicum codin
43	36.6	2.8	1380	21	AAZ34650	Human growth facto
44	36.6	2.8	2520	23	AAS73629	DNA encoding novel
45	36	2.7	1402	21	ABL60006	Proteinase A over r

rod repeats, H1 and H4 domains and a cysteine rich domain of a dystrophin gene -
Example 1; Page 52-53; 7lpp; English.

The present invention relates to an isolated nucleotide sequence encoding a dystrophin minigene. The minigene comprises N-terminal or modified N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control element. The dystrophin minigene in operable linkage with an expression control element, in a recombinant adeno-associated virus (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is human dystrophin DNA fragment encoding rods R23 and R24, hinge H4 and CR domain regions.

Sequence 1821 BP; 506 A; 451 C; 447 G; 417 T; 0 other;

Query Match 100.0%; Score 1321; DB 24; Length 1821;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	CGACTTCCAGCAGTTTCAGAACGACAGACGATGTACATAGGCGCTTCAAGAGGGAAATTGAA	60
DB	103	CGACTTCCAGCAGTTTCAGAACGACAGACGATGTACATAGGCGCTTCAAGAGGGAAATTGAA	162
QY	61	AACTAAAGAACCTGTAAATCATGAGTACTCTTGAGACTGTACGAATATTCTTGACAGAGCA	120
DB	163	AACTAAAGAACCTGTAAATCATGAGTACTCTTGAGACTGTACGAATATTCTTGACAGAGCA	222
QY	121	GCCTTTGGAAGGACTAGAGAACTCTACAGGAGCCAGAGAGCTGCCTCCTGAGGAGAG	180
DB	223	GCCTTTGGAAGGACTAGAGAACTCTACAGGAGCCAGAGAGCTGCCTCCTGAGGAGAG	282
QY	181	AGCCAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGA	240
DB	283	AGCCAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGA	342
QY	241	AAAATTGAACCTGCACCTCCGCTGACTGGCAGAGAGAAAATAGATGAGACCCCTTGAAGACT	300
DB	343	AAAATTGAACCTGCACCTCCGCTGACTGGCAGAGAGAAAATAGATGAGACCCCTTGAAGACT	402
QY	301	CCAGGAACCTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGGTGAT	360
DB	403	CCAGGAACCTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGGTGAT	462
QY	361	CAAGGGATCCTGGCAGCCCGTGGCGGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGA	420
DB	463	CAAGGGATCCTGGCAGCCCGTGGCGGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGA	522
QY	421	GAAAGTCAAGGCACCTTCGAGAGAAATTTGCGCCTCTGAAAGAGAACGTTGAGCCACGTCAA	480
DB	523	GAAAGTCAAGGCACCTTCGAGAGAAATTTGCGCCTCTGAAAGAGAACGTTGAGCCACGTCAA	582
QY	481	TGACCTTGCTCGCCAGCTTACCACCTTTGGGCATTCAGCTCTCACCGTATAACCTCAGCAC	540
DB	583	TGACCTTGCTCGCCAGCTTACCACCTTTGGGCATTCAGCTCTCACCGTATAACCTCAGCAC	642
QY	541	TCTGGAAGACCTGAACACCCAGATGGAAGCTTCTGAGGTGGCCGTCGAGGACCGAGTCAG	600
DB	643	TCTGGAAGACCTGAACACCCAGATGGAAGCTTCTGAGGTGGCCGTCGAGGACCGAGTCAG	702
QY	601	GCAGCTGCATGAAGCCACAGGGACTTTGGTCCAGCATCTCAGCACTTTCTTCCACGTC	660
DB	703	GCAGCTGCATGAAGCCACAGGGACTTTGGTCCAGCATCTCAGCACTTTCTTCCACGTC	762
QY	661	TGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCA	720
DB	763	TGTCCAGGGTCCCTGGGAGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCA	822
QY	721	CGAGACTCAAAACAACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGC	780

DB	823	CGAGACTCAAAACAACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCACTCTTTAGC	882
QY	781	TGACCTGAATAATGTAGATTTCTAGCTTATAGGACTGCCATGAAACTCCGAAAGACTGCA	840
DB	883	TGACCTGAATAATGTAGATTTCTAGCTTATAGGACTGCCATGAAACTCCGAAAGACTGCA	942
QY	841	GAAGGCCCTTTGCTTGATCTCTTGAGCCTGTCAGCTGTCATGTGATGCCCTTGACACGCA	900
DB	943	GAAGGCCCTTTGCTTGATCTCTTGAGCCTGTCAGCTGTCATGTGATGCCCTTGACACGCA	1002
QY	901	CAACCTCAAGCAAAATGACAGCCCATGGGATATCTTGCAGATTATTAATTGTTGACCCAC	960
DB	1003	CAACCTCAAGCAAAATGACAGCCCATGGGATATCTTGCAGATTATTAATTGTTGACCCAC	1062
QY	961	TATTTATGACCGCTTGAGCAAGAGCACAACAATTTGGTCAACGTCCTCTCTGCGTGA	1020
DB	1063	TATTTATGACCGCTTGAGCAAGAGCACAACAATTTGGTCAACGTCCTCTCTGCGTGA	1122
QY	1021	TATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAGGAGGATCCGCTGT	1080
DB	1123	TATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAGGAGGATCCGCTGT	1182
QY	1081	CCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAAGTACAG	1140
DB	1183	CCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAAGTACAG	1242
QY	1141	ATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCGCAGGCTGGGCCT	1200
DB	1243	ATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCGCAGGCTGGGCCT	1302
QY	1201	CCTTCTGTCATGATTTCTATCCAAATTTCCAAAGACAGTTGGGTGAAGTTGCATCCTTTGGGG	1260
DB	1303	CCTTCTGTCATGATTTCTATCCAAATTTCCAAAGACAGTTGGGTGAAGTTGCATCCTTTGGGG	1362
QY	1261	CAGTAACATTTGAGCCCAAGTGTCCCGGAGCTGCTCCAAATTTGCTAATAATAAGCCAGAGAT	1320
DB	1363	CAGTAACATTTGAGCCCAAGTGTCCCGGAGCTGCTCCAAATTTGCTAATAATAAGCCAGAGAT	1422
QY	1321	C 1321	
DB	1423	C 1423	

RESULT 2
AAD37232
ID AAD37232 standard; DNA; 2169 BP.
XX
AC AAD37232;
XX
DT 21-AUG-2002 (first entry)
XX
DE Human dystrophin rod, hinge and CR domain regions encoding DNA #1.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Homo sapiens.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
XX
PI Xiao X;
XX
DR WPI; 2002-049342/06.

XX New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -
XX
PS Example 1; Page 45-46; 71pp; English.
XX
CC The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is human
CC dystrophin DNA fragment encoding rods R22, R23 and R24, hinge H4 and
CC CR domain regions.
XX
SQ Sequence 2169 BP; 623 A; 529 C; 524 G; 493 T; 0 other;
Query Match 100.0%; Score 1321; DB 24; Length 2169;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CGACTTCCAGCAGTTTCAGAACGACGATGATACATAGGGCCCTCAAGAGGGAATTGAA 60
Db |||||
QY 451 CGACTTCCAGCAGTTTCAGAACGACGATGATACATAGGGCCCTCAAGAGGGAATTGAA 510
Db |||||
QY 61 AACTAAAGAACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCA 120
Db |||||
QY 511 AACTAAAGAACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCA 570
Db |||||
QY 121 GCCTTTGGAAGGACTAGAGAACTCTACCAGGAGGCCAGAGAGCTGCCTCCTGAGGAGAG 180
Db |||||
QY 571 GCCTTTGGAAGGACTAGAGAACTCTACCAGGAGGCCAGAGAGCTGCCTCCTGAGGAGAG 630
Db |||||
QY 181 AGCCAGAGATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGA 240
Db |||||
QY 631 AGCCAGAGATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGA 690
Db |||||
QY 241 AAAATTGAACCTGCACCTCCGCTGACTGGCAGAGAGAAAATAGATGAGACCCCTTGAAGACT 300
Db |||||
QY 691 AAAATTGAACCTGCACCTCCGCTGACTGGCAGAGAGAAAATAGATGAGACCCCTTGAAGACT 750
Db |||||
QY 301 CCAGGAACCTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGAT 360
Db |||||
QY 751 CCAGGAACCTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGAT 810
Db |||||
QY 361 CAAGGGATCCTGGCAGCCCGTGGCGATCTCCTCATTGACTCTCTCCAAGATCACCTCGA 420
Db |||||
QY 811 CAAGGGATCCTGGCAGCCCGTGGCGATCTCCTCATTGACTCTCTCCAAGATCACCTCGA 870
Db |||||
QY 421 GAAAGTCAAGGCACCTCGAGGAGAAATTGCGCCTCTGAAAGAGAGACGTGAGCCACGTCAA 480
Db |||||
QY 871 GAAAGTCAAGGCACCTCGAGGAGAAATTGCGCCTCTGAAAGAGAGACGTGAGCCACGTCAA 930
Db |||||
QY 481 TGACCTTGCTGCCAGCTTACCACTTTGGGCATTCAGCTCTCACCCTATACCTCAGCAC 540
Db |||||
QY 931 TGACCTTGCTGCCAGCTTACCACTTTGGGCATTCAGCTCTCACCCTATACCTCAGCAC 990
Db |||||
QY 541 TCTGGAAGACCTGAACACACAGATGGAAGCTTCTGCAGTGGCCGTCGAGGACCGAGTCAG 600
Db |||||
QY 991 TCTGGAAGACCTGAACACACAGATGGAAGCTTCTGCAGTGGCCGTCGAGGACCGAGTCAG 1050
Db |||||
QY 601 GCAGCTGCATGAAGCCACAGGGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGTC 660
Db |||||
QY 1051 GCAGCTGCATGAAGCCACAGGGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGTC 1110
Db |||||
QY 661 TGTCCAGGTCCTGGGAGAGAGCCATCTGCCAAACAAAGTGCCCTACTATATCAACCA 720
Db |||||

Db 1111 TGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCA 1170
QY 721 CGAGACTCAAAACAACCTTGTGGGACCACATCCAAAATGACAGAGCTCTACCAAGTCTTTAGC 780
Db |||||
QY 1171 CGAGACTCAAAACAACCTTGTGGGACCACATCCAAAATGACAGAGCTCTACCAAGTCTTTAGC 1230
Db |||||
QY 781 TGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCA 840
Db |||||
QY 1231 TGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCA 1290
Db |||||
QY 841 GAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTGAGCTGCATGTGATGCCTTGGACCAGCA 900
Db |||||
QY 1291 GAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTGAGCTGCATGTGATGCCTTGGACCAGCA 1350
Db |||||
QY 901 CAACCTCAAGCAAAATGACAGCCCATGGATATCCTGCAGATTATTAATTTGTTGACCAC 960
Db |||||
QY 1351 CAACCTCAAGCAAAATGACAGCCCATGGATATCCTGCAGATTATTAATTTGTTGACCAC 1410
Db |||||
QY 961 TATTTATGACCGCCTGGAGCAAGAGCACAAATTTGGTCAACGTCCCTCTCTGCGTGA 1020
Db |||||
QY 1411 TATTTATGACCGCCTGGAGCAAGAGCACAAATTTGGTCAACGTCCCTCTCTGCGTGA 1470
Db |||||
QY 1021 TATGTCTGAACTGGCTGCTGAATGTTTATGATACGGGACGAAACAGGAGGATCCGTGT 1080
Db |||||
QY 1471 TATGTCTGAACTGGCTGCTGAATGTTTATGATACGGGACGAAACAGGAGGATCCGTGT 1530
Db |||||
QY 1081 CCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAGCACATTTTGGAGCACAAAGTACAG 1140
Db |||||
QY 1531 CCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAGCACATTTTGGAGCACAAAGTACAG 1590
Db |||||
QY 1141 ATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAGCGAGGCTGGGCT 1200
Db |||||
QY 1591 ATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAGCGAGGCTGGGCT 1650
Db |||||
QY 1201 CCTTCTGCATGATTTATCCAAATTTCCAAAGACAGTTGGTGAAGTTCATCCTTTGGGG 1260
Db |||||
QY 1651 CCTTCTGCATGATTTATCCAAATTTCCAAAGACAGTTGGTGAAGTTCATCCTTTGGGG 1710
Db |||||
QY 1261 CAGTAACATTTGAGCCAAAGTGTCCGGAGCTGCTTCCAATTTGCTAATAAAGCCAGAGAT 1320
Db |||||
QY 1711 CAGTAACATTTGAGCCAAAGTGTCCGGAGCTGCTTCCAATTTGCTAATAAAGCCAGAGAT 1770
Db |||||
QY 1321 C 1321
Db 1771 C 1771
RESULT 3
AAD37240
ID AAD37240 standard; DNA; 3510 BP.
XX
AC AAD37240;
XX
DT 21-AUG-2002 (first entry)
XX
DE Human dystrophin minigene delta3510.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Homo sapiens.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
XX

Xiao X:
WPI; 2002-049342/06.
New dystrophin minigene for treating Duchenne or Becker muscular dystrophy comprises an N-terminal domain or modified N-terminal domain, rod repeats, H1 and H4 domains and a cysteine rich domain of a dystrophin gene.
Example 1; Page 51-52; 71pp; English.
The present invention relates to an isolated nucleotide sequence encoding a dystrophin minigene. The minigene comprises N-terminal or modified N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control element. The dystrophin minigene in operable linkage with an expression control element, in a recombinant adeno-associated virus or retrovirus is useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is human dystrophin minigene delta3510 containing nucleotides 1-1668 (N-terminus, hinge H1 and rods R1 and R2), 8407-10227 (rods R23 and R24, hinge H4 and CR domain) and 11047-11058 (dystrophin last 3 amino acids).
Sequence 3510 BP; 1073 A; 787 C; 828 G; 822 T; 0 other;
Query Match 100.0%; Score 1321; DB 24; Length 3510;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 1321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 CGACTTCCAGCAGTTTCAGAAAGCAGACGATGTACATAGGGCTTCAAGAGGGAATTGAA 60
1779 CGACTTCCAGCAGTTTCAGAAAGCAGACGATGTACATAGGGCTTCAAGAGGGAATTGAA 1838
61 AACTAAGAACCCTGTAAATCATGAGTACTCTCTGAGACTGTACGAATATTTCTGACAGAGCA 120
1839 AACTAAGAACCCTGTAAATCATGAGTACTCTCTGAGACTGTACGAATATTTCTGACAGAGCA 1898
121 GCCTTTGGAAGGACTAGAGAAACTCTACAGAGCCCGCAGAGAGCTGCCTCCTGAGGAGAG 180
1899 GCCTTTGGAAGGACTAGAGAAACTCTACAGAGCCCGCAGAGAGCTGCCTCCTGAGGAGAG 1958
181 AGCCCAAGAAATGTCACCTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGA 240
1959 AGCCCAAGAAATGTCACCTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGA 2018
241 AAAATTGAACCTGCATCCGCTGACTGGCAGAGAAATAAGATGAGACCCCTTGAAGAGACT 300
2019 AAAATTGAACCTGCATCCGCTGACTGGCAGAGAAATAAGATGAGACCCCTTGAAGAGACT 2078
301 CCAGGAACCTTCAAGAGGCGCACCGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGAT 360
2079 CCAGGAACCTTCAAGAGGCGCACCGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGAT 2138
361 CAAGGGATCCTGGCAGCCCGCTGGCGGATCTCCTCATTTGACTCTCTCCTCAAGATCACCTCGA 420
2139 CAAGGGATCCTGGCAGCCCGCTGGCGGATCTCCTCATTTGACTCTCTCCTCAAGATCACCTCGA 2198
421 GAAAGTCAAGGCACCTTCGAGGAGAAATTCGCGCTCTGAAAGAGAACTGAGCCACGTCAA 480
2199 GAAAGTCAAGGCACCTTCGAGGAGAAATTCGCGCTCTGAAAGAGAACTGAGCCACGTCAA 2258
481 TGACCTTGCTCGCCAGCTTACCACCTTGGGCGATTCAGCTCTCACCCTATACCTCAGCAC 540
2259 TGACCTTGCTCGCCAGCTTACCACCTTGGGCGATTCAGCTCTCACCCTATACCTCAGCAC 2318
541 TCTGGAAGACCTGAACACACAGATGGAAGCTTCTGAGGCTGGCCGTCGAGGACCGAGTCAG 600
2319 TCTGGAAGACCTGAACACACAGATGGAAGCTTCTGAGGCTGGCCGTCGAGGACCGAGTCAG 2378
601 GCAGCTGCATGAAGCCCAAGGAGCTTTGGTCCAGCATCTCAGCACTTTCTTCCACGTC 660

Db 2379 GCAGCTGCATGAAGCCCAAGGAGCTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGTC 2438
QY 661 TGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCA 720
Db 2439 TGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCA 2498
QY 721 CGAGACTCAAAACAACTTGTGGGACCATCCCAAAATGACAGAGCTCTACCACTCTTTAGC 780
Db 2499 CGAGACTCAAAACAACTTGTGGGACCATCCCAAAATGACAGAGCTCTACCACTCTTTAGC 2558
QY 781 TGACCTGAATAATGTAGATTTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCA 840
Db 2559 TGACCTGAATAATGTAGATTTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCA 2618
QY 841 GAAGGCCCTTGTGGATCTCTTGAGCCCTGTGATGATGATGATGATGATGATGATGATGATGAT 900
Db 2619 GAAGGCCCTTGTGGATCTCTTGAGCCCTGTGATGATGATGATGATGATGATGATGATGATGAT 2678
QY 901 CAACCTCAAGCAAAATGACAGCCCATGGATATCTTCAGATTTATTAATTTGTTGACCCAC 960
Db 2679 CAACCTCAAGCAAAATGACAGCCCATGGATATCTTCAGATTTATTAATTTGTTGACCCAC 2738
QY 961 TATTTATGACCCCTTGGAGCAAGAGCACAACAATTTGGTCAACGTCCTCTCTGCGTGA 1020
Db 2739 TATTTATGACCCCTTGGAGCAAGAGCACAACAATTTGGTCAACGTCCTCTCTGCGTGA 2798
QY 1021 TATGTGTCTGAACCTGGCTGCTGAATTTATGATACGGGACCAAGAGGATCCGCTGT 1080
Db 2799 TATGTGTCTGAACCTGGCTGCTGAATTTATGATACGGGACCAAGAGGATCCGCTGT 2858
QY 1081 CCTGTCTTTTAAACTGGCATCAATTTCCCTGTGTAAAGCACATTTTGAAGACAAAGTACAG 1140
Db 2859 CCTGTCTTTTAAACTGGCATCAATTTCCCTGTGTAAAGCACATTTTGAAGACAAAGTACAG 2918
QY 1141 ATACCTTTTCAAGCAAGTGCAAGTTTCAACAGGATTTTGTGACCAAGGCTGGGCT 1200
Db 2919 ATACCTTTTCAAGCAAGTGCAAGTTTCAACAGGATTTTGTGACCAAGGCTGGGCT 2978
QY 1201 CCTTCTGCATGATCTATCCAAATTTCCAAAGTTCAGAGAGTGGGTGAAGTTCATCCTTTGGGG 1260
Db 2979 CCTTCTGCATGATCTATCCAAATTTCCAAAGTTCAGAGAGTGGGTGAAGTTCATCCTTTGGGG 3038
QY 1261 CAGTAACATTTAGCCAAAGTGTCGGAGCTCCGAGCTGCTTCAATTTGCTAATAAAGCCAGAGAT 1320
Db 3039 CAGTAACATTTAGCCAAAGTGTCGGAGCTCCGAGCTGCTTCAATTTGCTAATAAAGCCAGAGAT 3098
QY 1321 C 1321
Db 3099 C 3099
RESULT 4
AAD37238
ID AAD37238 standard; DNA; 3531 BP.
XX
AC AAD37238;
XX 21-AUG-2002 (first entry)
XX Human dystrophin minigene delta3531.
DE Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
XX adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX Homo sapiens.
OS WO200183695-A2.
PN 08-NOV-2001.
XX 27-APR-2001; 2001WO-US13677.
PF
XX

PR 28-APR-2000; 2000US-200777P.
XX (XIAO/) XIAO X.
PA XIAO X;
XX WPI; 2002-049342/06.
XX
PT New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -
XX
PS Example 1; Page 50-51; 71pp; English.
XX
CC The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is human
CC dystrophin minigene delta3531 containing nucleotides 1-1341 (N-terminus,
CC hinge H1 and rod R1), 8059-10227 (rods R22, R23 and R24, hinge H4 and
CC CR domain) and 11047-11058 (dystrophin last 3 amino acids).
XX
SQ Sequence 3531 BP; 1071 A; 809 C; 824 G; 827 T; 0 other;

Query Match 100.0%; Score 1321; DB 24; Length 3531;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGACTTTCCAGCAGTTTCAGAAAGCAGAACGATGTACATAGGGCCCTCAAGAGGGAATTGAA 60
DB 1800 CGACTTTCCAGCAGTTTCAGAAAGCAGAACGATGTACATAGGGCCCTCAAGAGGGAATTGAA 1859

QY 61 AACTAAAGAACCTGTAAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCA 120
DB 1860 AACTAAAGAACCTGTAAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCA 1919

QY 121 GCCTTTGGAAGACTAGAGAAACTCTACCAGGAGCCCGCAGAGAGCTGCCTCCTGAGGAGAG 180
DB 1920 GCCTTTGGAAGACTAGAGAAACTCTACCAGGAGCCCGCAGAGAGCTGCCTCCTGAGGAGAG 1979

QY 181 AGCCAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGA 240
DB 1980 AGCCAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGA 2039

QY 241 AAAATTGAACCTGCACCTCCGCTGACTGGCAGAGAAATAATAGATGAGACCCCTTGAAAGACT 300
DB 2040 AAAATTGAACCTGCACCTCCGCTGACTGGCAGAGAAATAATAGATGAGACCCCTTGAAAGACT 2099

QY 301 CCAGGAACCTCAAGAGGCCCACGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGAT 360
DB 2100 CCAGGAACCTCAAGAGGCCCACGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGAT 2159

QY 361 CAAGGGATCCTGGCAGCCCGTGGCGATCTCCTCATTTGAGTCTCTCCAAAGATCACCTCGA 420
DB 2160 CAAGGGATCCTGGCAGCCCGTGGCGATCTCCTCATTTGAGTCTCTCCAAAGATCACCTCGA 2219

QY 421 GAAAGTCAAGGCACCTTCGAGGAGAAATTGCGCCTCTGAAAGAGAACGTGAGCCACGTCAA 480
DB 2220 GAAAGTCAAGGCACCTTCGAGGAGAAATTGCGCCTCTGAAAGAGAACGTGAGCCACGTCAA 2279

QY 481 TGACCTTGCTCGCCAGCTTACCAGCTTTGGGATTCAGCTCTCACCGTATACCTTCAGCAC 540
DB 2280 TGACCTTGCTCGCCAGCTTACCAGCTTTGGGATTCAGCTCTCACCGTATACCTTCAGCAC 2339

QY 541 TCTGGAAGACCTGAACACACAGATGGAAGCTTCTGCAGGTTGGCGCTCGAGGACCGAGTCAG 600
DB TCTGGAAGACCTGAACACACAGATGGAAGCTTCTGCAGGTTGGCGCTCGAGGACCGAGTCAG 600

Db 2340 TCTGGAAGACCTGAACACACAGATGGAAGCTTCTGCAGGTGGCGCTCGAGGACCGAGTCAG 2399
QY 601 GCAGCTGCATGAAGCCCCACAGGGACTTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGTC 660
Db 2400 GCAGCTGCATGAAGCCCCACAGGGACTTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGTC 2459
QY 661 TGTCCAGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCA 720
Db 2460 TGTCCAGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCA 2519
QY 721 CGAGACTCAAAACAACCTTGTCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGC 780
Db 2520 CGAGACTCAAAACAACCTTGTCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGC 2579
QY 781 TGACCTGAATAATGTCTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAAGACTGCA 840
Db 2580 TGACCTGAATAATGTCTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAAGACTGCA 2639
QY 841 GAAAGCCCTTTGCTTGGATCTCTTGAGCCCTGTCAGCTGCATGTGATGCCCTTGACCCAGCA 900
Db 2640 GAAAGCCCTTTGCTTGGATCTCTTGAGCCCTGTCAGCTGCATGTGATGCCCTTGACCCAGCA 2699
QY 901 CAACCTCAAGCAAAATGACCCAGCCCATGGATATCTCTGCAGATTATTAATTGTTGACCAC 960
Db 2700 CAACCTCAAGCAAAATGACCCAGCCCATGGATATCTCTGCAGATTATTAATTGTTGACCAC 2759
QY 961 TATTTATGACCGCCTGGAGCAAGAGCACAAATTTGGTCAACGTCCTCTCTGCGTGA 1020
Db 2760 TATTTATGACCGCCTGGAGCAAGAGCACAAATTTGGTCAACGTCCTCTCTGCGTGA 2819
QY 1021 TATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAAACAGGGAGGATCCCGTGT 1080
Db 2820 TATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAAACAGGGAGGATCCCGTGT 2879
QY 1081 CCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTTGGAGACACAAGTACAG 1140
Db 2880 CCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTTGGAGACACAAGTACAG 2939
QY 1141 ATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCGCAGGCTGGGCCT 1200
Db 2940 ATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCGCAGGCTGGGCCT 2999
QY 1201 CCTTCTGCATGATCTTATCCAAATTTCCAAAGACAGTTGGGTGAAGTTGCATCCTTTGGGG 1260
Db 3000 CCTTCTGCATGATCTTATCCAAATTTCCAAAGACAGTTGGGTGAAGTTGCATCCTTTGGGG 3059
QY 1261 CAGTACATTTGAGCCAAAGTGTCCGGAGCTGCTTCCAAATTTGCTAATAATAAGCCAGAGAT 1320
Db 3060 CAGTACATTTGAGCCAAAGTGTCCGGAGCTGCTTCCAAATTTGCTAATAATAAGCCAGAGAT 3119
QY 1321 C 1321
Db 3120 C 3120

RESULT 5
AAD37237
ID AAD37237 standard; DNA; 3858 BP.
XX
AC AAD37237;
XX
DT 21-AUG-2002 (first entry)
XX
DE Human dystrophin minigene delta3849.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
XX Becker muscular dystrophy; ds.
OS Homo sapiens.
XX
PN WO200183695-A2.
XX

100

100

100

100

OS Homo sapiens.
XX WO200183695-A2.
XX 08-NOV-2001.
XX 27-APR-2001; 2001WO-US13677.
XX 28-APR-2000; 2000US-200777P.
XX (XIAO/) XIAO X.
XX XIAO X;
XX WPI; 2002-049342/06.
XX
XX New dystrophin minigene for treating Duchenne or Becker muscular
XX dystrophy comprises an N-terminal domain or modified N-terminal domain,
XX rod repeats, H1 and H4 domains and a cysteine rich domain of a
XX dystrophin gene -
XX
XX Example 1; Page 46-47; 71pp; English.
XX
XX The present invention relates to an isolated nucleotide sequence encoding
XX a dystrophin minigene. The minigene comprises N-terminal or modified
XX N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
XX domains and cysteine-rich domains of dystrophin or utrophin genes. The
XX invention also relates to a recombinant adeno-associated virus (AAV)
XX comprising dystrophin minigene operably linked to an expression control
XX element. The dystrophin minigene in operable linkage with an expression
XX control element, in a recombinant adeno-associated virus or retrovirus is
XX useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
XX dystrophy (BMD) in a mammalian subject. The present sequence is human
XX dystrophin minigene delta3990 containing nucleotides 1-1668 (N-terminus,
XX hinge H1 and rods R1 and R2), 8059-10227 (rods R22, R23 and R24, hinge
XX H4 and CR domain) and 11047-11058 (dystrophin last 3 amino acids).
XX
XX Sequence 3999 BP; 1223 A; 907 C; 933 G; 936 T; 0 other;
XX
XX Query Match 100.0%; Score 1321; DB 24; Length 3999;
XX Best Local Similarity 100.0%; Pred. No. 0;
XX Matches 1321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 CGACTTTCCAGCAGTTTCAGAGCAGAACGATGTACATAGGGCCCTTCAAGAGGGGAATTGAA 60
XX |||||||
XX 2268 CGACTTTCCAGCAGTTTCAGAGCAGAACGATGTACATAGGGCCCTTCAAGAGGGGAATTGAA 2327
XX |||||||
XX 61 AACTAAAGAAACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCA 120
XX |||||||
XX 2328 AACTAAAGAAACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCA 2387
XX |||||||
XX 121 GCCTTTGGAAGGACTAGAGAACTCTACAGAGAGCCAGAGAGCTGCCCTCCTGAGGAGAG 180
XX |||||||
XX 2388 GCCTTTGGAAGGACTAGAGAACTCTACAGAGAGCCAGAGAGCTGCCCTCCTGAGGAGAG 2447
XX |||||||
XX 181 AGCCAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATFACTGAGTGGGA 240
XX |||||||
XX 2448 AGCCAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATFACTGAGTGGGA 2507
XX |||||||
XX 241 AAAATTGAACCTGCACCTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGACT 300
XX |||||||
XX 2508 AAAATTGAACCTGCACCTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGACT 2567
XX |||||||
XX 301 CCAGGAACCTTCAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGAT 360
XX |||||||
XX 2568 CCAGGAACCTTCAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGAT 2627
XX |||||||
XX 361 CAAGGGATCTTGGCAGCCCGCTGGCGGATCTCCTCAATGACTCTCTCCAAAGATCACCTCGA 420
XX |||||||
XX 2628 CAAGGGATCTTGGCAGCCCGCTGGCGGATCTCCTCAATGACTCTCTCCAAAGATCACCTCGA 2687
XX |||||||
XX 421 GAAAGTCAAGGCACCTTCGAGGAGAAATTTGGCCCTCTGAAAGAGAACGCTGAGCCACGTCAA 480
XX |||||||

Db 2688 GAAAGTCAAGGCACCTTCGAGGAGAAAATTGGCCCTCTGAAAGAGAACGTGAGCCACGTCAC 2747
QY
QY 481 TGACCTTGTCTGCGCAGCTTACCACCTTTGGGCATTCAGCTCTCACCGTATACCTCAGCAC 540
Db
Db 2748 TGACCTTGTCTGCGCAGCTTACCACCTTTGGGCATTCAGCTCTCACCGTATACCTCAGCAC 2807
QY
QY 541 TCTGGAAGACCTGAACACACACAGATGGAAGCTTCTGAGGTGGCCGTCGAGGACCGAGTCAG 600
Db
Db 2808 TCTGGAAGACCTGAACACACACAGATGGAAGCTTCTGAGGTGGCCGTCGAGGACCGAGTCAG 2867
QY
QY 601 GCAGCTGCATGAAGCCCAACAGGAGCTTTGGTCCAGCATCTCAGCAGCTTTCTTTCCAGCTC 660
Db
Db 2868 GCAGCTGCATGAAGCCCAACAGGAGCTTTGGTCCAGCATCTCAGCAGCTTTCTTTCCAGCTC 2927
QY
QY 661 TGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTCCCTACTATATCAACCA 720
Db
Db 2928 TGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTCCCTACTATATCAACCA 2987
QY
QY 721 CGAGACTCAAAACAACTTGTCTGGGAGAGCCATCCCAAAATGACAGAGCTCTACCAGTCTTAGC 780
Db
Db 2988 CGAGACTCAAAACAACTTGTCTGGGAGAGCCATCCCAAAATGACAGAGCTCTACCAGTCTTAGC 3047
QY
QY 781 TGACCTGAATAATGTCAGATTCCTCAGCTTATAGGACTGCCATGAAACTCCGAGAGACTGCA 840
Db
Db 3048 TGACCTGAATAATGTCAGATTCCTCAGCTTATAGGACTGCCATGAAACTCCGAGAGACTGCA 3107
QY
QY 841 GAAGGCCCTTTGCTTGGATCTCTTGAGCCCTGTGAGCTGCATGTGATGCCCTGGACCAGCA 900
Db
Db 3108 GAAGGCCCTTTGCTTGGATCTCTTGAGCCCTGTGAGCTGCATGTGATGCCCTGGACCAGCA 3167
QY
QY 901 CAACCTCAAGCAAAATGACACGAGAGCCATGATATCCTGCAGATTATTAATTTGACCCAC 960
Db
Db 3168 CAACCTCAAGCAAAATGACACGAGCCATGATATCCTGCAGATTATTAATTTGACCCAC 3227
QY
QY 961 TATTATGACCGCTGGAGCAAGAGCACAAATTTGGTCAAGTCCCTCTCTCGGTGGA 1020
Db
Db 3228 TATTATGACCGCTGGAGCAAGAGCACAAATTTGGTCAAGTCCCTCTCTCGGTGGA 3287
QY
QY 1021 TATGTGTCTGAACTGGCTGCTGAATGTTTATGATACGGGACGAGGAGGATCCGCTGT 1080
Db
Db 3288 TATGTGTCTGAACTGGCTGCTGAATGTTTATGATACGGGACGAGGAGGATCCGCTGT 3347
QY
QY 1081 CCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACACATTTGGAAGACAAAGTACAG 1140
Db
Db 3348 CCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACACATTTGGAAGACAAAGTACAG 3407
QY
QY 1141 ATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTGTGACGAGCGCAGGCTGGGCCT 1200
Db
Db 3408 ATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTGTGACGAGCGCAGGCTGGGCCT 3467
QY
QY 1201 CCTTCTGCATGATTTCTATCCAAATTCAGAGAGTTGGGTGAAGTTGCATCCTTTGGGG 1260
Db
Db 3468 CCTTCTGCATGATTTCTATCCAAATTCAGAGAGTTGGGTGAAGTTGCATCCTTTGGGG 3527
QY
QY 1261 CAGTAACATTTGAGCCCAAGTGTCCGAGCTGCTTCCAATTTGCTAATAAAGCCAGAGAT 1320
Db
Db 3528 CAGTAACATTTGAGCCCAAGTGTCCGAGCTGCTTCCAATTTGCTAATAAAGCCAGAGAT 3587
QY
QY 1321 C 1321
Db
Db 3588 C 3588

RESULT 7
AAD37230
ID AAD37230 standard; DNA; 4182 BP.
XX
AC AAD37230;
XX
DT 21-AUG-2002 (first entry)
XX
DE Human dystrophin minigene delta4173.
XX

Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin; adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD; Becker muscular dystrophy; ds.

OS Homo sapiens.
XX WO200183695-A2.
XX 08-NOV-2001.
XX 27-APR-2001; 2001WO-US13677.
XX 28-APR-2000; 2000US-200777P.
XX (XIAO/) XIAO X.

PI Xiao X;
XX WPI; 2002-049342/06.
XX

New dystrophin minigene for treating Duchenne or Becker muscular dystrophy comprises an N-terminal domain or modified N-terminal domain, rod repeats, H1 and H4 domains and a cysteine rich domain of a dystrophin gene.

Example 1; Page 43-44; 71pp; English.

The present invention relates to an isolated nucleotide sequence encoding a dystrophin minigene. The minigene comprises N-terminal or modified N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control element. The dystrophin minigene in operable linkage with an expression control element, in a recombinant adeno-associated virus or retrovirus is useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is human dystrophin minigene delta4173 containing nucleotides 1-1992 (N-terminus, hinge H1 and rods R1, R2 and R3), 8059-10227 (rods R2, R3 and R4, hinge H4 and CR domain) and 11047-11058 (dystrophin last 3 amino acids).

Sequence 4182 BP; 1309 A; 927 C; 970 G; 976 T; 0 other;

Query Match 100.0%; Score 1321; DB 24; Length 4182;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CGACTTTCCAGCAGTTTCAGAAAGCAGAACGATGTACATAGGGCCCTTCAAGAGGGAAATTGAA 60
2451 CGACTTTCCAGCAGTTTCAGAAAGCAGAACGATGTACATAGGGCCCTTCAAGAGGGAAATTGAA 2510
61 AACTAAAGAACCTGTATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCA 120
2511 AACTAAAGAACCTGTATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCA 2570
121 GCCTTTGGAAGGACTAGAGAAACTCTACCCAGGAGGCCAGAGAGCTGCCTCCTGAGGAGAG 180
2571 GCCTTTGGAAGGACTAGAGAAACTCTACCCAGGAGGCCAGAGAGCTGCCTCCTGAGGAGAG 2630
181 AGCCCAAGATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGAGGGA 240
2631 AGCCCAAGATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGAGGGA 2690
241 AAAATTGAACCTGCACCTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGAGACT 300
2691 AAAATTGAACCTGCACCTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGAGACT 2750
301 CCAGGAACCTTCAAGAGGCCACCGATGAGCTGGACCTCAAGCTCGGCCAAGCTGAGGTGAT 360
2751 CCAGGAACCTTCAAGAGGCCACCGATGAGCTGGACCTCAAGCTCGGCCAAGCTGAGGTGAT 2810
361 CAAGGATCCTGGCAGCCCGTGGCGGATCTCCTCATTTGACTCTCTCCCAAGATCACCTCGA 420

Db 2811 CAAGGGATCCTGGCAGCCCGTGGCGATCTCCTCATTTGACTCTCTCCAAAGATCACCTCGA 2870
QY 421 GAAAGTCAAGGCACCTTCGAGGAGAAATTTGCGCCTCTGAAAGAGAACTGAGCCACGTCAA 480
Db 2871 GAAAGTCAAGGCACCTTCGAGGAGAAATTTGCGCCTCTGAAAGAGAACTGAGCCACGTCAA 2930
QY 481 TGACCTTGCTCGCCAGCCTTACCACCTTTGGGCAATTCAGCTCTCACCGTATAAACCCTCAGCAC 540
Db 2931 TGACCTTGCTCGCCAGCCTTACCACCTTTGGGCAATTCAGCTCTCACCGTATAAACCCTCAGCAC 2990
QY 541 TCTGGAAGACCTGAACACACAGATGGAAGCTTTGTCAGGTGGCCGTCGAGGACCGAGTCAG 600
Db 2991 TCTGGAAGACCTGAACACACAGATGGAAGCTTTGTCAGGTGGCCGTCGAGGACCGAGTCAG 3050
QY 601 GCAGCTGCATGAAGCCACAGGGACTTTGGTCCAGCATCTCAGCAGCTTTCTTTCCACGTC 660
Db 3051 GCAGCTGCATGAAGCCACAGGGACTTTGGTCCAGCATCTCAGCAGCTTTCTTTCCACGTC 3110
QY 661 TGTCCAGGGTCCCTGGGAGAGAGCCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCA 720
Db 3111 TGTCCAGGGTCCCTGGGAGAGAGCCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCA 3170
QY 721 CGAGACTCAAAACAACTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAAGTCTTTAGC 780
Db 3171 CGAGACTCAAAACAACTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAAGTCTTTAGC 3230
QY 781 TGACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAAGACTGCA 840
Db 3231 TGACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAAGACTGCA 3290
QY 841 GAAGGCCCTTTGCTTGGATCTCTTGAGCCCTGTGAGCTGCATGTCATGTCCTTGGACCAAGCA 900
Db 3291 GAAGGCCCTTTGCTTGGATCTCTTGAGCCCTGTGAGCTGCATGTCATGTCCTTGGACCAAGCA 3350
QY 901 CAACCTCAAGCAAAATGACCAAGCATGATGATATCCTGCAGATTATTAATTGTTGACCAAC 960
Db 3351 CAACCTCAAGCAAAATGACCAAGCATGATGATATCCTGCAGATTATTAATTGTTGACCAAC 3410
QY 961 TATTATGACCGCCTGGAGCAAGAGACAAACAAATTTGGTCAACGTCCCTCTCTGCGTGA 1020
Db 3411 TATTATGACCGCCTGGAGCAAGAGACAAACAAATTTGGTCAACGTCCCTCTCTGCGTGA 3470
QY 1021 TATGTGTCTGAACCTGGCTGCTGAATGTTATGATACGGGACGAAACAGGGAGGATCCCGTGT 1080
Db 3471 TATGTGTCTGAACCTGGCTGCTGAATGTTATGATACGGGACGAAACAGGGAGGATCCCGTGT 3530
QY 1081 CCTGTCTTTTAAAACTGGCATCATTTCCCTGCTGTPAAAGCACATTTGGAAGACAAAGTACAG 1140
Db 3531 CCTGTCTTTTAAAACTGGCATCATTTCCCTGCTGTPAAAGCACATTTGGAAGACAAAGTACAG 3590
QY 1141 ATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTGTGACCGCAGCGAGGCTGGGCCT 1200
Db 3591 ATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTGTGACCGCAGCGAGGCTGGGCCT 3650
QY 1201 CCTTCTGCATGATTTCTATCCAAATTTCCAAAGACAGTTGGGTGAAGTTGCATCCTTTGGGG 1260
Db 3651 CCTTCTGCATGATTTCTATCCAAATTTCCAAAGACAGTTGGGTGAAGTTGCATCCTTTGGGG 3710
QY 1261 CAGTAACATTGAGCCCAAGTCTCCGGAGCTGCTTCCAAATTTGCTAAATAAAGCCAGAGAT 1320
Db 3711 CAGTAACATTGAGCCCAAGTCTCCGGAGCTGCTTCCAAATTTGCTAAATAAAGCCAGAGAT 3770
QY 1321 C 1321
Db 3771 C 3771
RESULT 8
AAD37259
ID AAD37259 standard; DNA; 4476 BP.
XX
AC AAD37259;
XX

DT 21-AUG-2002 (first entry)
XX
DE Adeno-associated virus vector plasmid, AAV-MCK-3510.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Unidentified.
XX
PN W0200183695-A2.
XX
XX
PD 08-NOV-2001.
XX
XX
PF 27-APR-2001; 2001WO-US13677.
XX
XX
PR 28-APR-2000; 2000US-200777P.
XX
XX
PA (XIAO/) XIAO X.
XX
XX XIAO X;
PI
XX
DR WPI; 2002-049342/06.
XX
XX
PT New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -
XX
PS Example 1; Page 63-65; 71pp; English.
XX
CC The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is AAV
CC vector plasmid construct containing human dystrophin minigenes, a muscle
CC creatine kinase (MCK) promoter and a small polyA signal sequence.
XX
SQ Sequence 4476 BP; 1252 A; 1096 C; 1127 G; 1001 T; 0 other;

Query Match
Best Local Similarity 100.0%; Score 1321; DB 24; Length 4476;
Matches 1321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGACTTTCCAGCAGTTCAGAACGACGATGTACATAGGGCCCTTCAAGAGGGAAATTGAA 60
Db 2535 CGACTTTCCAGCAGTTCAGAACGACGATGTACATAGGGCCCTTCAAGAGGGAAATTGAA 2594
QY 61 AACTAAAGAACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCA 120
Db 2595 AACTAAAGAACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCA 2654
QY 121 GCCTTTTGGAGGACTAGAGAACTCTACCAGGAGCCGACAGAGCTGCCTCCTGAGGAGAG 180
Db 2655 GCCTTTTGGAGGACTAGAGAACTCTACCAGGAGCCGACAGAGCTGCCTCCTGAGGAGAG 2714
QY 181 AGCCAGATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGA 240
Db 2715 AGCCAGATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGA 2774
QY 241 AAAATTGAACCTGCACCTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGACT 300
Db 2775 AAAATTGAACCTGCACCTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGACT 2834
QY 301 CCAGGAACCTTCAAGAGGCCGAGTGGACCTCAAGCTGCGCCCAAGCTGAGGTGAT 360
Db 3601 CCGAGGAACCTTCAAGAGGCCGAGTGGACCTCAAGCTGCGCCCAAGCTGAGGTGAT 3654

Db 2835 CCAGGAACCTTCAAGAGGCCGAGTGGACCTCAAGCTGCGCCCAAGCTGAGGTGAT 2894
QY 361 CAAGGGATCCTGGCAGCCCCGCTGGGGGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGA 420
Db 2895 CAAGGGATCCTGGCAGCCCCGCTGGGGGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGA 2954
QY 421 GAAAGTCAAGGCACTTCGAGGAGAAATTCGCGCTCTGAAAAGAGACGTGAGCCACGTCAA 480
Db 2955 GAAAGTCAAGGCACTTCGAGGAGAAATTCGCGCTCTGAAAAGAGACGTGAGCCACGTCAA 3014
QY 481 TGACCTTGTCTCGCCAGCTTACCACCTTTGGGCAATTCAGCTCTCACCCTATAACCTCAGCAC 540
Db 3015 TGACCTTGTCTCGCCAGCTTACCACCTTTGGGCAATTCAGCTCTCACCCTATAACCTCAGCAC 3074
QY 541 TCTGGAAGACCTGAACACCCACAGATGGAAGCTTCTGAGGTGGCCGTCGAGGACCGAGTCAG 600
Db 3075 TCTGGAAGACCTGAACACCCACAGATGGAAGCTTCTGAGGTGGCCGTCGAGGACCGAGTCAG 3134
QY 601 GCAGCTGCATGAAGCCACAGGAGCTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGTC 660
Db 3135 GCAGCTGCATGAAGCCACAGGAGCTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGTC 3194
QY 661 TGTCCAGGCTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCA 720
Db 3195 TGTCCAGGCTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCA 3254
QY 721 CGAGACTCAAAACAACTTGTCTGGGAGAGCCATCCCAAAATGACAGAGCTCTACCAGTCTTAGC 780
Db 3255 CGAGACTCAAAACAACTTGTCTGGGAGAGCCATCCCAAAATGACAGAGCTCTACCAGTCTTAGC 3314
QY 781 TGACCTGAATAATGTCTCAGATCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCA 840
Db 3315 TGACCTGAATAATGTCTCAGATCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCA 3374
QY 841 GAAGGCCCTTTGTCTGGATCTCTTGAGCCTGTGAGCTGATGATGATGATGATGATGATGATGAT 900
Db 3375 GAAGGCCCTTTGTCTGGATCTCTTGAGCCTGTGAGCTGATGATGATGATGATGATGATGATGAT 900
QY 901 CAACCTCAAGCAAAATGACCAAGAGCCATGGATATCCCTGAGATTTAATTGTTTGACCA 960
Db 3435 CAACCTCAAGCAAAATGACCAAGAGCCATGGATATCCCTGAGATTTAATTGTTTGACCA 3494
QY 961 TATTTATGACCCGCTGGAGCAAGAGCAACAATTTGGTCAACGTCCTCTCTCGGTGGA 1020
Db 3495 TATTTATGACCCGCTGGAGCAAGAGCAACAATTTGGTCAACGTCCTCTCTCGGTGGA 3554
QY 1021 TATGTCTTGAAGTGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080
Db 3555 TATGTCTTGAAGTGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 3614
QY 1081 CCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTTGAAGACAAAGTACAG 1140
Db 3615 CCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTTGAAGACAAAGTACAG 3674
QY 1141 ATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAAGGCTGGGCT 1200
Db 3675 ATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAAGGCTGGGCT 3734
QY 1201 CCTTCTGCATGATCTATCCAAATTCGAAGACAGTTGGGTGAAGTTGCATCCTTTGGGG 1260
Db 3735 CCTTCTGCATGATCTATCCAAATTCGAAGACAGTTGGGTGAAGTTGCATCCTTTGGGG 3794
QY 1261 CAGTAACATGAGCCCAAGTGTCCGAGCTGCTTCCAATTTGCTAATAAGCCAGAGAT 1320
Db 3795 CAGTAACATGAGCCCAAGTGTCCGAGCTGCTTCCAATTTGCTAATAAGCCAGAGAT 3854
QY 1321 C 1321
Db 3855 C 3855

RESULT 9
AAD37258

AAD37258 standard; DNA; 4498 BP.
 AAD37258;
 21-AUG-2002 (first entry)
 Adeno-associated virus vector plasmid, AAV-MCK-3531.
 Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
 adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
 Becker muscular dystrophy; ds.
 Chimeric - Homo sapiens.
 Chimeric - Unidentified.
 WO200183695-A2.
 08-NOV-2001.
 27-APR-2001; 2001WO-US13677.
 28-APR-2000; 2000US-200777P.
 (XIAO/) XIAO X.
 Xiao X;
 WPI; 2002-049342/06.
 New dystrophin minigene for treating Duchenne or Becker muscular
 dystrophy comprises an N-terminal domain or modified N-terminal domain,
 rod repeats, H1 and H4 domains and a cysteine rich domain of a
 dystrophin gene -
 Example 1; page 62-63; 71pp; English.
 The present invention relates to an isolated nucleotide sequence encoding
 a dystrophin minigene. The minigene comprises N-terminal or modified
 N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
 domains and cysteine-rich domains of dystrophin or utrophin genes. The
 invention also relates to a recombinant adeno-associated virus (AAV)
 comprising dystrophin minigene operably linked to an expression control
 element. The dystrophin minigene in operable linkage with an expression
 control element, in a recombinant adeno-associated virus or retrovirus is
 useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
 dystrophy (BMD) in a mammalian subject. The present sequence is AAV
 vector plasmid construct containing human dystrophin minigenes, a muscle
 creatine kinase (MCK) promoter and a small polyA signal sequence.
 Sequence 4498 BP; 1251 A; 1118 C; 1123 G; 1006 T; 0 other;
 Query Match 100.0%; Score 1321; DB 24; Length 4498;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 CGACTTTCAGCAGTTCAGAAAGCAGAACCGATGTACATAGGGCCTTCAAGAGGGGAATTGAA 60
 Db 2557 CGACTTTCAGCAGTTCAGAAAGCAGAACCGATGTACATAGGGCCTTCAAGAGGGGAATTGAA 2616
 QY 61 AACTAAAGAACCTGTAAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCA 120
 Db 2617 AACTAAAGAACCTGTAAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCA 2676
 QY 121 GCCTTTGGAAAGGACTAGAGAAACTCTACCAGAGGCCAGAGAGCTGCCTCCTGAGGAGAG 180
 Db 2677 GCCTTTGGAAAGGACTAGAGAAACTCTACCAGAGGCCAGAGAGCTGCCTCCTGAGGAGAG 2736
 QY 181 AGCCCAGAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGA 240
 Db 2737 AGCCCAGAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGA 2796
 QY 241 AAAATTGAACCTGCACCTCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGAGACT 300


```

RESULT 10
AAD37257
ID   AAD37257 standard; DNA; 4825 BP.
XX
AC   AAD37257;
XX
DT   21-AUG-2002 (first entry)
XX
DE   Adeno-associated virus vector plasmid, AAV-MCK-delta3849.
XX
KW   Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW   adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW   Becker muscular dystrophy; ds.
XX
OS   Chimeric - Homo sapiens.
OS   Chimeric - Unidentified.
XX
PN   WO200183695-A2.
XX
PD   08-NOV-2001.
XX
PF   27-APR-2001; 2001WO-US13677.
XX
PR   28-APR-2000; 2000US-200777p.
XX
PA   (XIAO/) XIAO X.
XX
PI   Xiao X;
XX
DR   WPI; 2002-049342/06.
XX
PT   New dystrophin minigene for treating Duchenne or Becker muscular
PT   dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT   rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT   dystrophin gene -
XX
PS   Example 1; Page 61-62; 71pp; English.
XX
CC   The present invention relates to an isolated nucleotide sequence encoding
CC   a dystrophin minigene. The minigene comprises N-terminal or modified
CC   N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC   domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC   invention also relates to a recombinant adeno-associated virus (AAV)
CC   comprising dystrophin minigene operably linked to an expression control
CC   element. The dystrophin minigene in operable linkage with an expression
CC   control element, in a recombinant adeno-associated virus or retrovirus is
CC   useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC   dystrophy (BMD) in a mammalian subject. The present sequence is AAV
CC   vector plasmid construct containing human dystrophin minigenes, a muscle
CC   creatine kinase (MCK) promoter and a small polyA signal sequence.
XX
SQ   Sequence 4825 BP; 1369 A; 1175 C; 1204 G; 1077 T; 0 other;

Query Match
Best Local Similarity 100.0%; Score 1321; DB 24; Length 4825;
Matches 1321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY   1   CGACTTTCAGCAGTTCAGAGCAGAACGATGTACATAGGGCTTCAAGAGGGGAATTGAA 60
Db    2884 CGACTTTCAGCAGTTCAGAGCAGAACGATGTACATAGGGCTTCAAGAGGGGAATTGAA 60
QY   61   AACTAAAGAACCTGTATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCA 120
Db    2944 AACTAAAGAACCTGTATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCA 120
QY   121   GCCTTTGGAAGGACTAGAGAACTCTACCAGGAGCCCCAGAGAGCTGCCTCTCTGAGGAGAG 180
Db    3004 GCCTTTGGAAGGACTAGAGAACTCTACCAGGAGCCCCAGAGAGCTGCCTCTCTGAGGAGAG 180
QY   181   AGCCCGAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGA 240
Db    181 AGCCCGAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGA 240

```


Mon Apr 28 09:28:22 2003

1321 C 1321
4204 C 4204

RESULT 11
AAD37263
ID AAD37263 standard; DNA; 4848 BP.
XX
AC AAD37263;
XX
DT 21-AUG-2002 (first entry)
XX
DE Adeno-associated virus (AAV) vector plasmid, AAV-CMV-delta3849.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Cytomegalovirus.
OS Chimeric - Unidentified.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
XX
PI Xiao X;
XX
DR WPI; 2002-049342/06.
XX
PT New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -
XX
PS Example 1; Page 68-70; 71pp; English.
XX
CC The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus (DMD) and Becker muscular
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is AAV
CC vector plasmid construct containing human dystrophin minigenes, a
CC cytomagalovirus (CMV) promoter and a small polyA signal sequence.
XX
SQ Sequence 4848 BP; 1405 A; 1144 C; 1180 G; 1119 T; 0 other;

Query Match
Best Local Similarity 100.0%; Score 1321; DB 24; Length 4848;
Matches 1321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 CGACTTTCCAGCAGTTTCCAGAGCAGAACGATGATAGGCGCTTCAAGAGGGAATTGAA 60
2907 CGACTTTCCAGCAGTTTCCAGAGCAGAACGATGATAGGCGCTTCAAGAGGGAATTGAA 2966
61 AACTAAGAGAACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCA 120
2967 AACTAAGAGAACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCA 3026
121 GCCTTTGGAAGGAGCTAGAGAAACTCTACCAGGAGGCCACAGAGAGCTGCCTCCTGAGGAGAG 180

3027 GCCTTTGGAAGGAGCTAGAGAAACTCTACCAGGAGGCCACAGAGAGCTGCCTCCTGAGGAGAG 3086
181 AGCCGAGAAATGTCACCTGGCTTCTACGAAAGCAGGCTGAGGAGCTCAATAGTGGGA 240
3087 AGCCGAGAAATGTCACCTGGCTTCTACGAAAGCAGGCTGAGGAGCTCAATAGTGGGA 3146
241 AAAATTGAACCTGCACCTCCGCTGACTGGCAGAGAAAATAGATGAGACCCCTTGAAAGACT 300
3147 AAAATTGAACCTGCACCTCCGCTGACTGGCAGAGAAAATAGATGAGACCCCTTGAAAGACT 3206
301 CCAGGAACCTCAAGAGGCCACCGGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGGTGAT 360
3207 CCAGGAACCTCAAGAGGCCACCGGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGGTGAT 3266
361 CAAGGGATCCTGGCAGCCCGTGGCGGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGA 420
3267 CAAGGGATCCTGGCAGCCCGTGGCGGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGA 3326
421 GAAAGTCAAGGCACCTTCGAGGAGAGAAATTCGCGCTCTCTGAAAGAGAACGTCAGCCACGTC 480
3327 GAAAGTCAAGGCACCTTCGAGGAGAGAAATTCGCGCTCTCTGAAAGAGAACGTCAGCCACGTC 3386
481 TGACCTTGCTCGCCAGCTTACCACCTTTGGGCAATTCAGCTCTCACCCGTATAAACCCTCAGCAC 540
3387 TGACCTTGCTCGCCAGCTTACCACCTTTGGGCAATTCAGCTCTCACCCGTATAAACCCTCAGCAC 3446
541 TCTGGAAGACCTGAACACACAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAG 600
3447 TCTGGAAGACCTGAACACACAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAG 3506
601 GCAGCTGCATGAAGCCACAGGCACTTTGGTCCAGCATCTCAGCACTTTCTTCCACGTC 660
3507 GCAGCTGCATGAAGCCACAGGCACTTTGGTCCAGCATCTCAGCACTTTCTTCCACGTC 3566
661 TGTCAGAGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCTACTATATCAACCA 720
3567 TGTCAGAGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCTACTATATCAACCA 3626
721 CGAGACTCAAAACAACCTTGTGGGACCATTCCTCCAAATGACAGAGCTCTACCAAGTCTTTAGC 780
3627 CGAGACTCAAAACAACCTTGTGGGACCATTCCTCCAAATGACAGAGCTCTACCAAGTCTTTAGC 3686
781 TGACCTGAATAATGTCAGATTTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCA 840
3687 TGACCTGAATAATGTCAGATTTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCA 3746
841 GAAGGCCCTTTGCTTGGATCTCTTGGACCTGTGACCTGTGATGCTTGGACAGCA 900
3747 GAAGGCCCTTTGCTTGGATCTCTTGGACCTGTGACCTGTGATGCTTGGACAGCA 3806
901 CAACCTCAAGCAAAATGACAGCCCATGGATATCCTGCGAGATTTAATTTGTTGACCAC 960
3807 CAACCTCAAGCAAAATGACAGCCCATGGATATCCTGCGAGATTTAATTTGTTGACCAC 3866
961 TATTTATGACCGCTGGAGCAAGAGACACAAATTTGGTCAACGTCCTCTCTCGTGGA 1020
3867 TATTTATGACCGCTGGAGCAAGAGACACAAATTTGGTCAACGTCCTCTCTCGTGGA 3926
1021 TATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGAGGATCCGTGT 1080
3927 TATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGAGGATCCGTGT 3986
1081 CCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTTGGAGAGCAAGTACAG 1140
3987 CCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTTGGAGAGCAAGTACAG 4046
1141 ATACCTTTTCAAGCAAGTGGCAAGTTTCAACAGGATTTTGTGACCGAGGCTGGGCTT 1200
4047 ATACCTTTTCAAGCAAGTGGCAAGTTTCAACAGGATTTTGTGACCGAGGCTGGGCTT 4106
1201 CCTTCTGATGATTTCTATCCAAATTTCCAAGACAGTTTGGTGAAGTTGTCATCCTTTGGGG 1260

Db 4107 CCTTCTGCATGATCTATCCAAATTCAGACAGATTGGTGAAGTTGCATCTTTGGGG 4166
QY 1261 CAGTAACATTGAGCCAAAGTGTCGGAGCTGCTTCCAATTTGCTAATAATAAGCCAGAGAT 1320
Db 4167 CAGTAACATTGAGCCAAAGTGTCGGAGCTGCTTCCAATTTGCTAATAATAAGCCAGAGAT 4226
QY 1321 C 1321
Db 4227 C 4227

RESULT 12
AAD37256
ID AAD37256 standard; DNA; 4966 BP.
XX
AC AAD37256;
XX
DT 21-AUG-2002 (first entry)
XX
DE Adeno-associated virus vector plasmid, AAV-MCK-delta3990.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX

OS Chimeric - Homo sapiens.
OS Chimeric - Unidentified.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
XX
PI Xiao X;
XX
DR WPI; 2002-049342/06.
XX
PT New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -
XX
PS Example 1; Page 59-60; 71pp; English.

CC The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is AAV
CC vector plasmid construct containing human dystrophin minigenes, a muscle
CC creatine kinase (MCK) promoter and a small polyA signal sequence.

XX
SQ Sequence 4966 BP; 1403 A; 1216 C; 1232 G; 1115 T; 0 other;
Query Match 100.0%; Score 1321; DB 24; Length 4966;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGACTTCCAGCAGTTTCAGAAAGCAGACGATGTACATAGGCGCTTCAAGAGGGAATTGAA 60
Db 3025 CGACTTCCAGCAGTTTCAGAAAGCAGACGATGTACATAGGCGCTTCAAGAGGGAATTGAA 3084
QY 61 AACTAAAGAACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTCTGACAGAGCA 120

Db 3085 AACTAAAGAACCTGTAATCATGAGTACTCTGAGACTGTACGAATATTCTGACAGAGCA 3144
QY 121 GCCTTTGGAAGGACTAGAGAAACTCTACAGGAGCCAGAGAGCTGCCTCCTGAGGAGAG 180
Db 3145 GCCTTTGGAAGGACTAGAGAAACTCTACAGGAGCCAGAGAGCTGCCTCCTGAGGAGAG 3204
QY 181 AGCCACAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGA 240
Db 3205 AGCCACAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGA 3264
QY 241 AAAATTGAACCTGCACCTCCGCTGACTGGCAGAGAGAAAATAGATGAGACCTTTGAAAGACT 300
Db 3265 AAAATTGAACCTGCACCTCCGCTGACTGGCAGAGAGAAAATAGATGAGACCTTTGAAAGACT 3324
QY 301 CCAGGAACTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCCCAAGCTGAGGTGAT 360
Db 3325 CCAGGAACTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCCCAAGCTGAGGTGAT 3384
QY 361 CAAGGGATCCTGGCAGCCCGTGGCGGATCTCCTCATTGACTCTCTCCAAAGATCACCTCGA 420
Db 3385 CAAGGGATCCTGGCAGCCCGTGGCGGATCTCCTCATTGACTCTCTCCAAAGATCACCTCGA 3444
QY 421 GAAAGTCAAGGCACCTTCGAGGAGAGAAATTGGCGCTCTGAAAGAGAACGTGAGCCACGTCAA 480
Db 3445 GAAAGTCAAGGCACCTTCGAGGAGAGAAATTGGCGCTCTGAAAGAGAACGTGAGCCACGTCAA 3504
QY 481 TGACCTTGCTCGCCAGCTTACCACCTTTGGGCAATTCAGCTCTCACCGTATAACCTCAGCAC 540
Db 3505 TGACCTTGCTCGCCAGCTTACCACCTTTGGGCAATTCAGCTCTCACCGTATAACCTCAGCAC 3564
QY 541 TCTGGAAGACCTGAACACACAGATGGAAGCTTTCGAGGTGGCGCTCGAGGACCCGAGTCAG 600
Db 3565 TCTGGAAGACCTGAACACACAGATGGAAGCTTTCGAGGTGGCGCTCGAGGACCCGAGTCAG 3624
QY 601 GCAGCTGCATGAAGCCCAACAGGAGCTTTGGTCCAGCATCTCAGCACCTTTCTTTCCACGTC 660
Db 3625 GCAGCTGCATGAAGCCCAACAGGAGCTTTGGTCCAGCATCTCAGCACCTTTCTTTCCACGTC 3684
QY 661 TGTCCAGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCA 720
Db 3685 TGTCCAGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCA 3744
QY 721 CGAGACTCAAAACAACTTGTGGGACCATCCCAAAATGACAGAGCTCTACCACTCTTAGC 780
Db 3745 CGAGACTCAAAACAACTTGTGGGACCATCCCAAAATGACAGAGCTCTACCACTCTTAGC 3804
QY 781 TGACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAACCTCCGAAGACTGCA 840
Db 3805 TGACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAACCTCCGAAGACTGCA 3864
QY 841 GAAGGCCCTTTGCTTGGATCTCTGAGCCTGTGAGCTGTCATGTGATGCTTGGACCAAGCA 900
Db 3865 GAAGGCCCTTTGCTTGGATCTCTGAGCCTGTGAGCTGTCATGTGATGCTTGGACCAAGCA 3924
QY 901 CAACCTCAAGCAAAATGACAGCCCATGGATATCCTGCAGATTATTAAATTGTTGACCAC 960
Db 3925 CAACCTCAAGCAAAATGACAGCCCATGGATATCCTGCAGATTATTAAATTGTTGACCAC 3984
QY 961 TATTATGACCGCTGGAGCAAGAGCACAACAAATTTGGTCAACGTCCCTCTCTCGGTGGA 1020
Db 3985 TATTATGACCGCTGGAGCAAGAGCACAACAAATTTGGTCAACGTCCCTCTCTCGGTGGA 4044
QY 1021 TATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGAGGATCCGPT 1080
Db 4045 TATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGAGGATCCGPT 4104
QY 1081 CCTGTCTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAGTACAG 1140
Db 4105 CCTGTCTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAGTACAG 4164
QY 1141 ATACCTTTTCAAGCAAGTGGCAAGTTCACAGAGATTGTTGACCAAGCGAGGCTGGCCT 1200

Db 4165 ATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAGCGCAGGCTGGGCCT 4224
QY 1201 CCTTCTGCATGATTTCTATCCAAATCCCAAGACAGTTGGGTGAAGTTGCATCCTTTGGGGG 1260
Db 4225 CCTTCTGCATGATTTCTATCCAAATCCCAAGACAGTTGGGTGAAGTTGCATCCTTTGGGGG 4284
QY 1261 CAGTAACATTGAGCCCAAGTGTCCGGAGCTGCTTCCAATTTGCTAATAATAAGCCAGAGAT 1320
Db 4285 CAGTAACATTGAGCCCAAGTGTCCGGAGCTGCTTCCAATTTGCTAATAATAAGCCAGAGAT 4344
QY 1321 C 1321
Db 4345 C 4345

RESULT 13
AAD37262
ID AAD37262 standard; DNA; 4990 BP.
XX
AC AAD37262;
DT 21-AUG-2002 (first entry)
XX
DE Adeno-associated virus (AAV) vector plasmid, AAV-CMV-delta3990.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Cytomegalovirus.
OS Chimeric - Unidentified.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
XX
PI Xiao X;
XX
DR WPI; 2002-049342/06.
XX
PT New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -
XX
PS Example 1; Page 67-68; 71pp; English.
XX
CC The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is AAV
CC vector plasmid construct containing human dystrophin minigenes, a
CC cytomagalovirus (CMV) promoter and a small polyA signal sequence.

XX
SQ Sequence 4990 BP; 1439 A; 1185 C; 1208 G; 1158 T; 0 other;
Query Match 100.0%; Score 1321; DB 24; Length 4990;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 1321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CGACTTTCCAGCAGTTTCAAGACGAGAACGATGTACATAGGCGCTTCAAGAGGGAATTGAA 60
Db 3049 CGACTTTCCAGCAGTTTCAAGACGAGAACGATGTACATAGGCGCTTCAAGAGGGAATTGAA 3108
QY 61 AACTAAAGAACCTGTAAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCA 120
Db 3109 AACTAAAGAACCTGTAAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCA 3168
QY 121 GCCTTTGGAAGGACTAGAGAAACTCTTACCAGGAGGCCAGAGAGTGGCTCCTGAGGAGAG 180
Db 3169 GCCTTTGGAAGGACTAGAGAAACTCTTACCAGGAGGCCAGAGAGTGGCTCCTGAGGAGAG 3228
QY 181 AGCCAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGA 240
Db 3229 AGCCAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGA 3288
QY 241 AAAATTGAACCTGCACTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGAGACT 300
Db 3289 AAAATTGAACCTGCACTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGAGACT 3348
QY 301 CCAGGAACCTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGGCCAAGCTGAGGTGAT 360
Db 3349 CCAGGAACCTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGGCCAAGCTGAGGTGAT 3408
QY 361 CAAGGGATCCTGGCAGCCCCGCTGGGGGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGA 420
Db 3409 CAAGGGATCCTGGCAGCCCCGCTGGGGGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGA 3468
QY 421 GAAAGTCAAGGCACCTTCGAGGAGAAAAATTTGGCCCTCTGAAAAGAGAACGTGAGCCACGTCAA 480
Db 3469 GAAAGTCAAGGCACCTTCGAGGAGAAAAATTTGGCCCTCTGAAAAGAGAACGTGAGCCACGTCAA 3528
QY 481 TGACCTTTGCTCGCCAGCTTACCACCTTTGGGGCATTTACCTCTCACCCTGATAACCTCAGCAC 540
Db 3529 TGACCTTTGCTCGCCAGCTTACCACCTTTGGGGCATTTACCTCTCACCCTGATAACCTCAGCAC 3588
QY 541 TCTGGAAGACCTGAACACACAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAG 600
Db 3589 TCTGGAAGACCTGAACACACAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAG 3648
QY 601 GCAGCTGCATGAAGCCCAACAGGACCTTTGGTCCAGCATCTCAGACATTTCTTCCACGTC 660
Db 3649 GCAGCTGCATGAAGCCCAACAGGACCTTTGGTCCAGCATCTCAGACATTTCTTCCACGTC 3708
QY 661 TGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCA 720
Db 3709 TGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCA 3768
QY 721 CGAGACTCAAAACAACTTGTGGGACCATCCCAAAATGACAGAGCTCTACAGCTCTTTAGC 780
Db 3769 CGAGACTCAAAACAACTTGTGGGACCATCCCAAAATGACAGAGCTCTACAGCTCTTTAGC 3828
QY 781 TGACCTGAATAATGTCAGATTTCTCAGCTTATAGGACTGCCATGAAACTCCGAAAGACTGCA 840
Db 3829 TGACCTGAATAATGTCAGATTTCTCAGCTTATAGGACTGCCATGAAACTCCGAAAGACTGCA 3888
QY 841 GAAGGCCCTTTGCTTGGATCTCTTGAGCCCTGTCAGCTGCATGTGATGCCCTTGGACAGCA 900
Db 3889 GAAGGCCCTTTGCTTGGATCTCTTGAGCCCTGTCAGCTGCATGTGATGCCCTTGGACAGCA 3948
QY 901 CAACCTCAAGCAAAATGACCAGCCCATGGATATCTCTGCAGATTATTAATTGTTTGACCAC 960
Db 3949 CAACCTCAAGCAAAATGACCAGCCCATGGATATCTCTGCAGATTATTAATTGTTTGACCAC 4008
QY 961 TATTTATGACCGCTGGAGCAAGAGCACAAATTTGGTCAACGTCCCTCTCTGCGTGGGA 1020
Db 4009 TATTTATGACCGCTGGAGCAAGAGCACAAATTTGGTCAACGTCCCTCTCTGCGTGGGA 4068
QY 1021 TATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGACGAACAGGGAGGATCCGTGT 1080
Db 4069 TATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGACGAACAGGGAGGATCCGTGT 4128
QY 1081 CCTGTCTTTTAAACCTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAG 1140

QY	458	AAAGAGAACGTGAGCCACGTCAATGACCTTGCTCGCCAGCTTACCACCTTTGGGCATTCAG	517
Db	2172	AAAGAGAACGTGAGCCACGTCAATGACCTTGCTCGCCAGCTTACCACCTTTGGGCATTCAG	2231
QY	518	CTCTCACCGTATAACCTTCAGCACTCTGGAAGACCTGAACACCAGATGGAAGCTTCTGCAG	577
Db	2232	CTCTCACCGTATAACCTTCAGCACTCTGGAAGACCTGAACACCAGATGGAAGCTTCTGCAG	2291
QY	578	GTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGAAGCCCACAGGGACTTTGGTCCAGCA	637
Db	2292	GTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGAAGCCCACAGGGACTTTGGTCCAGCA	2351
QY	638	TCTCAGCACCTTTCTTTCCACGCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAAC	697
Db	2352	TCTCAGCACCTTTCTTTCCACGCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAAC	2411
QY	698	AAAGTCCCCTACTATATCAACCACGAGACTCAAAACAACCTTGCTGGGACCATCCCAAAATG	757
Db	2412	AAAGTCCCCTACTATATCAACCACGAGACTCAAAACAACCTTGCTGGGACCATCCCAAAATG	2471
QY	758	ACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACT	817
Db	2472	ACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACT	2531
QY	818	GCCATGAACCTCCGAAGACTGCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTACGCT	877
Db	2532	GCCATGAACCTCCGAAGACTGCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTACGCT	2591
QY	878	GCATGTGATGCCCTTGACCCAGCACAACTCAAGCAAAAATGACCAGCCCATGGATATCCTG	937
Db	2592	GCATGTGATGCCCTTGACCCAGCACAACTCAAGCAAAAATGACCAGCCCATGGATATCCTG	2651
QY	938	CAGATTATTAAATTGTTGACCACTATTTATGACCCGCTGGAGCAAGAGCACAAATTTG	997
Db	2652	CAGATTATTAAATTGTTGACCACTATTTATGACCCGCTGGAGCAAGAGCACAAATTTG	2711
QY	998	GTCAACGTCCCCTCTCTGCGTGGATATGTCTGAACTGGCTGCTGAATGTTTATGATACG	1057
Db	2712	GTCAACGTCCCCTCTCTGCGTGGATATGTCTGAACTGGCTGCTGAATGTTTATGATACG	2771
QY	1058	GGACGAACAGGGAGGATCCGTGTCCTGTCTTTTAAAACTGGCATCATTTCCCTGTGTAAA	1117
Db	2772	GGACGAACAGGGAGGATCCGTGTCCTGTCTTTTAAAACTGGCATCATTTCCCTGTGTAAA	2831
QY	1118	GCACATTGGAAGACAAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCACAGGATTT	1177
Db	2832	GCACATTGGAAGACAAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCACAGGATTT	2891
QY	1178	TGTGACCAAGCGCAGGCTGGGCCCTCCTTCTGATGATTTCTATCCAAATTCCAAGACAGTTG	1237
Db	2892	TGTGACCAAGCGCAGGCTGGGCCCTCCTTCTGATGATTTCTATCCAAATTCCAAGACAGTTG	2951
QY	1238	GGTGAAGTTGCATCCTTTTGGGGGCAGTAACAATGAGCCAAGTGTCCGGAGCTGCTTCCAA	1297
Db	2952	GGTGAAGTTGCATCCTTTTGGGGGCAGTAACAATGAGCCAAGTGTCCGGAGCTGCTTCCAA	3011
QY	1298	TTTGGCTAATAATAAGCCAGAGATC	1321
Db	3012	TTTGGCTAATAATAAGCCAGAGATC	3035

Search completed: April 25, 2003, 08:43:55
Job time : 247.366 secs

```

Db      2898 GCCAGAGATC 2907
|||||
RESULT 15
AAD37242
ID      AAD37242 standard; DNA; 3446 BP.
XX
XX
AC      AAD37242;
XX
XX      21-AUG-2002 (first entry)
XX
XX      Human dystrophin minigene delta3447.
DE
XX      Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW      adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW      Becker muscular dystrophy; ds.
XX
XX      Homo sapiens.
OS
XX      WO200183695-A2.
PN
XX
XX      08-NOV-2001.
PD
XX      27-APR-2001; 2001WO-US13677.
PF
XX      28-APR-2000; 2000US-200777P.
PR
XX      (XIAO/) XIAO X.
PA
XX
XX      Xiao X;
PI
XX
XX      WPI; 2002-049342/06.
DR
XX
XX      New dystrophin minigene for treating Duchenne or Becker muscular
PT      dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT      rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT      dystrophin gene -
XX
XX      Example 1; Page 53-54; 71pp; English.
PS
XX
XX      The present invention relates to an isolated nucleotide sequence encoding
CC      a dystrophin minigene. The minigene comprises N-terminal or modified
CC      N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC      domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC      invention also relates to a recombinant adeno-associated virus (AAV)
CC      comprising dystrophin minigene operably linked to an expression control
CC      element. The dystrophin minigene in operable linkage with an expression
CC      control element, in a recombinant adeno-associated virus (DMD) and Becker muscular
CC      dystrophy (BMD) in a mammalian subject. The present sequence is human
CC      dystrophin minigene delta3447 containing nucleotides 1-1992 (N-terminus,
CC      hinge H1 and rods R1, R2 and R3), 8749-10227 (rod R24, hinge H4 and
CC      CR domain) and 11047-11058 (dystrophin last 3 amino acids).
XX
XX      Sequence 3446 BP; 1074 A; 766 C; 787 G; 819 T; 0 other;
SQ

Query Match      78.5%; Score 1037.6; DB 24; Length 3446;
Best Local Similarity 99.6%; Pred. No. 8.4e-311;
Matches 1040; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY      278 ATAGATGAGACCCCTTGAAGACTCCAGGAACCTCAAGAGGCCACGGATGAGCTGGACCTC 337
      ||| |||||
Db      1992 ACAGCACAGACCCCTTGAAGACTCCAGGAACCTCAAGAGGCCACGGATGAGCTGGACCTC 2051

QY      338 AAGCTGCGCCCAAGCTGAGTGATCAAGGGATCCTGGCAGCCCCGTGGCGATCTCCTCATT 397
      |||||
Db      2052 AAGCTGCGCCCAAGCTGAGTGATCAAGGGATCCTGGCAGCCCCGTGGCGATCTCCTCATT 2111

QY      398 GACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGCACTTCGAGGAGAAATTCGCGCTCTG 457
      |||||
Db      2112 GACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGCACTTCGAGGAGAAATTCGCGCTCTG 2171

```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 07:58:12 ; Search time 1246 Seconds
(without alignments)
17170.339 Million cell updates/sec

Title: US-09-845-416-10_COPY_1800_3120

Perfect score: 1321

Sequence: 1 cgactttccagcaggttcaga.....ctaataagccagagatc 1321

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308016

Minimum DB seq length: 0

Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST: *

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: gb_gss:*
- 18: em_gss_hum:*
- 19: em_gss_inv:*
- 20: em_gss_pln:*
- 21: em_gss_vrt:*
- 22: em_gss_fun:*
- 23: em_gss_mam:*
- 24: em_gss_mus:*
- 25: em_gss_other:*
- 26: em_gss_pro:*
- 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

result No.	Score	Query Match	Length	ID	Description
1	604	45.7	757	9	AL556247
2	593	44.9	620	14	BQ640063
3	540.6	40.9	2135	11	AK013510
4	538.2	40.7	770	12	BG719710
5	502.8	38.1	652	10	BB629984
6	491.6	37.2	704	10	BB610411

7	479.6	36.3	854	9	AI196693
8	432.8	32.8	898	12	BF182065
9	410.4	31.1	599	10	BB666688
10	385.2	29.2	684	9	AL641565
11	381.8	28.9	423	9	AA460476
12	380.4	28.8	633	9	AA460476
13	358.2	27.1	665	10	BE201973
14	330	25.0	608	9	AL672616
15	329	24.9	532	10	BE334408
16	325.6	24.6	532	13	BJ036615
17	315.4	23.9	541	13	BI546771
18	310	23.5	488	13	BM488542
19	297.8	22.5	597	9	AL795652
20	293.8	22.2	696	13	BJ075057
21	281.8	21.3	573	13	BI446561
22	276.6	20.9	647	9	AL652775
23	273.2	20.7	417	12	BF387048
24	269.2	20.4	726	10	BB627285
25	265.8	20.1	784	12	BG212445
26	256.4	19.4	733	12	BF144271
27	250	18.9	502	9	AL602076
28	247.6	18.7	501	9	AA771140
29	242.8	18.4	835	13	BI553820
30	239.8	18.2	475	12	BG207912
31	221	16.7	663	13	BM488464
32	220.4	16.7	834	13	BI729851
33	206.4	15.6	479	14	BQ304046
34	203	15.4	810	12	BE911128
35	198	15.0	714	14	BM931093
36	196.2	14.9	579	9	AL121550
37	190.8	14.4	230	14	N87737
38	184.8	14.0	307	12	BF226333
39	176.8	13.4	750	13	BI730168
40	173	13.1	1002	12	BE798394
41	153.4	11.6	481	10	BE557463
42	140	10.6	585	10	AV963180
43	134.6	10.2	770	9	AA392372
44	133.2	10.1	170	9	AA095279
45	130.2	9.9	663	10	BB083824

ALIGNMENTS

RESULT 1
AL556247
LOCUS
DEFINITION
AL556247 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DK001YB17 5
prime, mRNA sequence.
ACCESSION
AL556247
VERSION
AL556247.1 GI:12898746
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 757)
AUTHORS
Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished (2001)
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. .757
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DK001YB17"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 190 a 197 c 188 g 182 t

Query Match 45.7%; Score 604; DB 9; Length 757;
Best Local Similarity 100.0%; Pred. No. 2.5e-170;
Matches 604; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

2Y 718 CCACGAGACTCAAAACAACTTGCTGGGACCATCCAAAATGACAGAGCTCTACCACTCTTT 777
128 CCACGAGACTCAAAACAACTTGCTGGGACCATCCAAAATGACAGAGCTCTACCACTCTTT 187
2Y 778 AGCTGACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT 837
188 AGCTGACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT 247
2Y 838 GCAGAAAGGCCCTTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGGACCA 897
248 GCAGAAAGGCCCTTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGGACCA 307
898 GCACAACTCAAGCAAAATGACCAAGCAAGAGCAACAATTTGGTCAACGCTCCCTCTCTGCGT 957
308 GCACAACTCAAGCAAAATGACCAAGCAAGAGCAACAATTTGGTCAACGCTCCCTCTCTGCGT 367
958 CACTATTATGACCGCCTGGAGCAAGAGCAACAATTTGGTCAACGCTCCCTCTCTGCGT 1017
368 CACTATTATGACCGCCTGGAGCAAGAGCAACAATTTGGTCAACGCTCCCTCTCTGCGT 427
1018 GGATATGTGCTGAAGTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCG 1077
428 GGATATGTGCTGAAGTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCG 487
1078 TGTCCCTGCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTA 1137
488 TGTCCCTGCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTA 547
1138 CAGATACCTTTTCAAGCAAGTGGCAAGTTTCAACAGGATTTGTGACCAAGCGCAGGCTGGG 1197
548 CAGATACCTTTTCAAGCAAGTGGCAAGTTTCAACAGGATTTGTGACCAAGCGCAGGCTGGG 607
1198 CCTCCTTCTGCATGATCTATCCAAATTCGAAGACAGTTGGGTGAAGTTGCATCCTTTGG 1257
608 CCTCCTTCTGCATGATCTATCCAAATTCGAAGACAGTTGGGTGAAGTTGCATCCTTTGG 667
1258 GGGCAGTAACATTGAGCAAGTGTCCGGAGCTGCTTCCAAATTTGCTAATAAAGCCAGA 1317
668 GGGCAGTAACATTGAGCAAGTGTCCGGAGCTGCTTCCAAATTTGCTAATAAAGCCAGA 727
1318 GATC 1321
728 GATC 731

RESULT 2
BQ640063
LOCUS BQ640063 620 bp mRNA linear EST 15-JUL-2002
DEFINITION he23g04.y1 Human Retina cDNA (Un-normalized, unamplified): hd/he
Homo sapiens cDNA clone he23g04 5', mRNA sequence.

ACCESSION BQ640063
VERSION BQ640063.1 GI:21764522
KEYWORDS EST.

SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 620)

AUTHORS Wistow,G., Bernstein,S.L., Wyatt,M.K., Ray,S., Behal,A., Touchman ,J.W., Bouffard,G., Smith,D. and Peterson,K.
TITLE Expressed sequence tag analysis of human retina for the NEIBank Project: Retbinding, an abundant, novel retinal cDNA and alternative splicing of other retina-preferred gene transcripts
JOURNAL Mol. Vis. 8 (4), (2002) In press
COMMENT Contact: Wistow G
Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 23 row: g column: 04
Seq primer: M13RPl reverse primer (ABI).

FEATURES
source Location/Qualifiers
1..620
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="he23g04"
/clone_lib="Human Retina cDNA (Un-normalized, unamplified)": hd/he"
/tissue_type="Retina"
/dev_stage="Adult"
/lab_host="EMDH10B"
/note="Organ: Eye; Vector: pSPORT1; Neural retina tissue was dissected from two 80 year old donors with no observed eye disease. 100ug of total RNA was used for library construction. A directionally cloned cDNA library in the pSPORT1 vector (Life Technologies) was constructed at Bioserve Biotechnology (Laurel MD) essentially following the protocols of the SuperScript Plasmid System full details of which are contained in the manufacturer's instruction manual (http://www.lifetech.com/). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGACTAGTCTAGATCGGAGCGGCGCC(T)15-3']. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

BASE COUNT 165 a 148 c 144 g 163 t
ORIGIN

Query Match 44.9%; Score 593; DB 14; Length 620;
Best Local Similarity 100.0%; Pred. No. 4.4e-167;
Matches 593; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 729 AAACAACCTGCTGGGACCATCCCAAAATGACAGAGCTCTTACCACTCTTTAGCTGACCTGA 788
Db 1 AAACAACCTGCTGGGACCATCCCAAAATGACAGAGCTCTTACCACTCTTTAGCTGACCTGA 60
QY 789 ATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAACTCCGAAGACTGCAGAAGGCC 848
Db 61 ATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAACTCCGAAGACTGCAGAAGGCC 120
QY 849 TTTGCTTGGATCTCTTGAGCCTGTGAGCTGCATGTGATGCTTGGACCAACCACTCA 908
Db 121 TTTGCTTGGATCTCTTGAGCCTGTGAGCTGCATGTGATGCTTGGACCAACCACTCA 180
QY 909 AGCAAAATGACCAAGCCCATGGATATCCTGCAGATTATTAATTGTTGACCACATATTTATG 968
Db 181 AGCAAAATGACCAAGCCCATGGATATCCTGCAGATTATTAATTGTTGACCACATATTTATG 240
QY 969 ACCGCCTGGAGCAAGACGACACAATTTGGTCAACGTCCTCTCTCGTGGATATGTGTC 1028
Db 241 ACCGCCTGGAGCAAGACGACACAATTTGGTCAACGTCCTCTCTCGTGGATATGTGTC 300
QY 1029 TGAAGTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCGTGTCTCTT 1088
Db 301 TGAAGTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCGTGTCTCTT 360
QY 1089 TTAAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAGATACCTTT 1148
Db 361 TTAAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAGATACCTTT 420

ORIGIN

Query Match 40.9%; Score 540.6; DB 11; Length 2135;
Best Local Similarity 93.5%; Pred. No. 6.9e-151;
Matches 564; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 718 CCACGAGACTCAAAACAACCTTGCTGGGACCACCCAAAATGACAGAGCTCTACCAGTCTTT 777
Db 144 CCACGAGACCCAAACACCTTGTGGGACCACCCAAAATGACAGAGCTCTACCAGTCTTT 203
QY 778 AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT 837
Db 204 AGCTGACCTGAATAATGTTCAGGTTCTCCGGGTATAGGACTGCCATGAAGCTCAGAAGCT 263
QY 838 GCAGAGAGGCCCTTTGCTTGGATCTCTTGGCCCTGTCAGCTGCATGTGATGCCCTTGGACCA 897
Db 264 CCAGAGAGGCCCTTTGCTTGGATCTCTTGGCCCTGTCAGCTGCATGTGATGCCCTTGGACCA 323
QY 898 GCACAACCTCAAGCAAAATGACCAAGAGCAGCAACAATTTGGTCAACGCTCCCTCTCTCGCT 957
Db 324 GCACAACCTCAAGCAAAATGACCAAGAGCAGCAACAATTTGGTCAACGCTCCCTCTCTCGCT 383
QY 958 CACTATTTATGACCGCTGGAGCAAGAGCAGCAACAATTTGGTCAACGCTCCCTCTCTCGCT 1017
Db 384 TACAATTTATGATCGTCTGGAGCAAGAGCAGCAACAATTTGGTCAACGCTCCCTCTCTCGCT 443
QY 1018 GGATATGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACCAACAGGAGGATCCG 1077
Db 444 GGATATGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACCAACAGGAGGATCCG 503
QY 1078 TGTCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCAGCATTTGGAAGACAAGTA 1137
Db 504 TGTCTGTCTTTTAAACTGGCATCATTTCTCTGTGTAAAGCAGCATTTGGAAGACAAGTA 563
QY 1138 CAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCGAGGCTGGG 1197
Db 564 CAGATACCTTTTCAAGCAAGTGGCAAGTTCAACTGGCTTTTGTGACCGAGGCTGGG 623
QY 1198 CCTCCTCTGCTGATGATCTATCCAAATCCAAAGACAGTTGGGTGAAGTTGCATCCTTTGG 1257
Db 624 TCTTCTCTGCTGATGATCTATTCAAATCCCAAGACAGTTGGGTGAAGTTGCTTCTCTTGG 683
QY 1258 GGGCAGTAACATTGAGCAAGTGTCCGGAGCTGCTTCCAAATTTGCTAATAATAAGCCAGA 1317
Db 684 GGGCAGTAACATTGAGCCGAGTGTGAGGAGCTGCTTCCAAATTTGCCAATAATAAACCTGA 743
QY 1318 GAT 1320
Db 744 GAT 746

RESULT 4

BG719710 770 bp mRNA linear EST 08-MAY-2001
LOCUS 602690430F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4822807 5',
DEFINITION mRNA sequence.

ACCESSION BG719710
VERSION BG719710.1 GI:13998897
KEYWORDS EST.
SOURCE human.

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS NIH-MGC <http://mgi.nci.nih.gov/>.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLAM10731 row: 1 column: 08
High quality sequence stop: 767.

FEATURES

source

Location/Qualifiers
1..770
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4822807"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',
size-selected for average insert size 2.2 kb and
normalized to ROT 5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI/NIH), National
Institutes of Health). Note: this is a NIH_MGC Library."

BASE COUNT 189 a 199 c 198 g 184 t

ORIGIN

Query Match 40.7%; Score 538.2; DB 12; Length 770;
Best Local Similarity 96.5%; Pred. No. 1.7e-150;
Matches 572; Conservative 0; Mismatches 18; Indels 3; Gaps 2;

QY 718 CCACGAGACTCAAAACAACCTTGCTGGGACCACCCAAAATGACAGAGCTCTACCAGTCTTT 777
Db 150 CCACGAGACTCAAAACAACCTTGCTGGGACCACCCAAAATGACAGAGCTCTACCAGTCTTT 209
QY 778 AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT 837
Db 210 AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT 269
QY 838 GCAGAGAGGCCCTTTGCTTGGATCTCTTGGCCCTGTCAGCTGCATGTGATGCCCTTGGACCA 897
Db 270 GCAGAGAGGCCCTTTGCTTGGATCTCTTGGCCCTGTCAGCTGCATGTGATGCCCTTGGACCA 329
QY 898 GCACAA--CCTCAAGCAAAATGACCAAGCCCATGGATATCCTGCAGATTATTAATTGTTG 955
Db 330 GCACAAATCCTCAAGTCAGATCAGATGACCAAGCCCATGGATATCCTGCAGATTATTAATTGTTG 389
QY 956 ACCACTATTTATGACCGCCTGGAGCAAGAGCACAACAATTTGGTCAACGTCCTCTCTGC 1015
Db 390 ACCACTATTTATGACCGCCTGGAGCAAGAGCACAACAATTTGGTCAACGTCCTCTCTGC 449
QY 1016 GTGGATATGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACCAACAGGAGGATC 1075
Db 450 GTGGATATGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACCAACAGGAGGATC 509
QY 1076 CGTGCTCTGCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCAGATTGGAAGACAAG 1135
Db 510 CGTGCTCTGCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCAGATTGGAAGACAAG 569
QY 1136 TACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTGTTGTGACCGAGGCTG 1195
Db 570 TACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTGTTGTGACCGAGGCTG 628
QY 1196 GGCCTCCTCTGCTGATGATCTATCCAAATCCAAAGACAGTTGGGTGAAGTTGCATCCTTT 1255
Db 629 GGCCTCCTCTGCTGATGATCTATCCAAATCCAAAGACAGTGGGTGAAGTTGCATCCTTT 688
QY 1256 GGGGGCAGTAACATTGAGCCAAAGTGTCCGGAGCTGCTTCCAATTTGCTAATAA 1308
Db 689 GGGGGCAGTAACATAGAGCCAAAGTGTCCGGAGCTGTTCCAATTTGCTAATAA 741

RESULT 5

BB629984
LOCUS

BB629984 652 bp mRNA linear EST 31-AUG-2001
LOCUS

,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H., Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T., Muramatsu,M. and Hayashizaki,Y.
RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
Unpublished (2001)
Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp/
URL:http://genome.gsc.riken.go.jp/
Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara,Y. and Hayashizaki,Y.
Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

e mouse tissues.
Location/Qualifiers
1. .704
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="1810074E05"
/clone_lib="RIKEN full-length enriched, 10 day old male pancreas"
/sex="male"
/tissue_type="pancreas"
/dev_stage="10 day old"
/note="Vector: pSPORT1; Site_1: SalI; Site_2: NotI; This clone is among a rearranged set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos , and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dt)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the X chromosome, 1998, Hum Mol Genet 7: 1967-1978."
Mol Genet 7: 1967-1978."
173 a 185 c 167 g 179 t
37.2%; Score 491.6; DB 10; Length 704;

	Best Local Similarity	93.0%;	Pred. No. 1.8e-136;						
	Matches	515;	Conservative	0;	Mismatches	39;	Indels	0;	Gaps
QY	718	CCACGAGACTCAAACAACCTGCTGGGACCACATCCCAAAATGACAGAGCTCTACCACTCTTT	777						
Db	137	CCACGAGACCCCAACCACTTGTGGGACCACCCCAAAATGACAGAGCTCTACCACTCTTT	196						
QY	778	AGCTGACCTGAATAATGTACATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	837						
Db	197	AGCTGACCTGAATAATGTACATTCTCAGCTTATAGGACTGCCATGAAAGCTCAGAAGGCT	256						
QY	838	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTGAGCTGCATGTGATGCCTTGGACCA	897						
Db	257	CCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTGAGCTGCATGTGATGCCTTGGACCA	316						
QY	898	GCACAACCTCAAGCAAAATGACCAAGCCATGGATATCCTGCAGATTATTAATTGTTGAC	957						
Db	317	GCACAACCTCAAGCAAAATGACCAAGCCATGGATATCCTGCAGATAATTAACCTGTTGAC	376						
QY	958	CACATATTATGACCGCCTGGAGCAAGAGACACAACAATTTGGTCAACGTCCCTCTCTGCGT	1017						
Db	377	TACAATTATGATCGTCTGGAGCAAGAGACACAACAATCTGGTCAATGTCCCTCTCTGTGT	436						
QY	1018	GGATATGTGTCTGAACCTGGCTGCTGAATGTTATGATACAGGACGAGGAGGATCCG	1077						
Db	437	GGATATGTGTCTCAACTGGCTTCTCAATGTTATGATACAGGACGAGGAGGATCCG	496						
QY	1078	TGCTCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTA	1137						
Db	497	TGCTCTGTCTTTTAAACTGGCATCATTTCTCTGTGTAAAGCACATTTGGAAGACAAGTA	556						
QY	1138	CAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACGAGCGAGGCTGGG	1197						
Db	557	CAGATACCTTTTCAAGCAAGTGGCAAGTTCAACTGGCTTTTGTGACCAACGTAAGCTGGG	616						
QY	1198	CTCTCTTCTGTCATGATTTCTATCCAAATTTCCAAAGACAGTGGGTGAAGTTGCATCCTTGG	1257						
Db	617	TCTTCTTCTGTCATGATTTCTATCCAAATTTCCAAAGACAGTGGGTGAAGTTGCATCCTTGG	676						
QY	1258	GGCAGTAACATTG 1271							
Db	677	GGCAGTAACATTG 690							

RESULT 7
AI196693
LOCUS
DEFINITION
AI196693
854 bp
mRNA
linear
EST 14-OCT-1998
ui53e10.y1 Sugano mouse liver mlia Mus musculus cDNA clone
IMAGE:1886154 5' similar to gb:M18533 DYSTROPHIN (HUMAN); gb:M68859
Mouse dystrophin mRNA, exons 1-7 and complete cds (MOUSE);, mRNA
sequence.
AI196693
AI196693.1 GI:3749299
EST.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 854)
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
Waterston,R.
The WashU-HHMI Mouse EST Project
Unpublished (1996)
Contact: Marra M/Mouse EST Project
WashU-HHMI Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@watson.wustl.edu

TITLE
JOURNAL
COMMENT

BASE COUNT
ORIGIN

Query Match

718	QY	CCACGAGACTCAAAACAACCTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTT	777
129	Db	CCACGAGACCCAAACCACTTGTGGGACCACCCCAAAATGACAGAGCTCTACCAGTCTTT	
778	QY	AGCTGACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	837
189	Db	AGCTGACCTGAATAATGTCAGTTCTCCGCGTATAGGACTGCCATGAACTCAGAAGGCT	248
838	QY	GCAGAAGGCCCTTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCA - TGTGATGCCCTTGGACC	896
249	Db	CCAGAAGGCCCTTGTCTTGGATCTCTTGAGCCTGTCAGCTGCAGTGTGATGCCCTGGACC	307
897	QY	AGCACAACTCAAGCAAAATGA - CCAGCCCCATGGATATCCTGCAGATTATAA - TTGTTT	954
308	Db	AGCACAACTCAAGCAAAATGAGCCAGCCCCATGGATATCCTGCAGATAATTAACGTGTTT	367
955	QY	GACCACATTTATGACCGCCTGGAGCAAGACACAA - CAATTTGGTCAACGTCCCTCTCT	1011
368	bb	GACTACAATTTATGATCGTCTGGAGCAAGAGCACAAAGCAATCTGGTCAATGTCCCTCTCT	427


```
QY 898 GCACAACTCAAGCAAAATGACCCATGGATATCCTGCAGATTATTAATTGTTGAC 957
|||||
Ddb 207 GCACAACTCAAGCAAAATGACCCATGGATATCCTGCAGATTATTAATTGTTGAC 266
|||||
QY 958 CACTATTATGACCGCTGGAGCAAGAGCACAACAATTGGTCAACGTCCTCTCTGCGT 1017
|||||
Ddb 267 CACTATTATGACCGCTGGAGCAAGAGCACAACAATTGGTCAACGTCCTCTCTGCGT 326
|||||
QY 1018 GGATATGTGTCTGAACCTGGCTGCTGAAT-GTTTATGATACGGGACGAACAGGGAGGATCC 1076
|||||
Ddb 327 GGATATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCC 386
|||||
QY 1077 GTGTCCTGTCTTTTAAACTGGCATCATTTCCCTGTG 1113
|||||
Ddb 387 GTGTCCTGTCTTTTAAACCGGCATCATTTCCCTGTG 423
|||||

RESULT 12
LOCUS AL796733 633 bp mRNA linear EST 27-JUN-2002
DEFINITION AL796733 XGC-neurula Silurana tropicalis cDNA clone TNeul43c03 5',
mRNA sequence.
ACCESSION AL796733.1 GI:21582437
VERSION AL796733
KEYWORDS EST.
SOURCE western clawed frog.
ORGANISM Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Silurana.
1 (bases 1 to 633)
Taylor, R., Ashurst, J.L., Croning, M.D.R., Zorn, A.M. and Rogers, J.
Sanger Xenopus tropicalis EST project 2002
Unpublished (2001)
Contact: Taylor R
Sanger Centre
Hinxtion, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TNeul43c03.plcSP6
Sequencing primer: PlcSP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
Location/Qualifiers
1..633
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="TNeul43c03"
/dev_stage="neurula"
/lab_host="Escherichia coli DH10B"
/note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dt primed from 5ug of poly A+ RNA from neurula.
EcoRI-NotI cut cDNA was then ligated into pCS107 with
EcoRI at the 5' end and NotI at the 3' end."
BASE COUNT 169 a 154 c 151 g 159 t
ORIGIN
Query Match 28.8%; Score 380.4; DB 9; Length 633;
Best Local Similarity 84.3%; Pred. No. 5.3e-103;
Matches 440; Conservative 0; Mismatches 81; Indels 1; Gaps 1;

QY 718 CCACGAGACTCAACAACTTGCTGGGACCATCCAAAATGACAGAGCTCTACCACTTTT 777
|||||
Ddb 113 CCATGAGACAAACCACTGCTGGGATCATCCAAAATGACAGATTATACCAATCTTT 172
|||||
QY 778 AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT 837
|||||
Ddb 173 ATCTGACCTGAACAATGTGCGATTCTCATCATACAGAACTGCCATGAAGCTAAAGAGATT 232
|||||
QY 838 GCAGAGGCCCTTTGCTGGATCTCTTGAGCCCTGTGACCTGTCATGTGATGCCTTGGACCA 897
|||||
```

```
Db 233 GCAAAAGGCCTTGTGTGGATTGCTAAGGCTGTCTGCAGCTTGTGAAGCCTTGGACCA 292
QY 898 GCACAACTCAAGCAAAATGACCCATGGATATCCTGCAGATTATTAATTGTTGAC 957
|||||
Ddb 293 GCACAACTGAAGCAGAATGACCCAGTGTGATGGACATCCTGCAGATTATTAATTGCTTGAC 352
|||||
QY 958 CACTATTATGACCGCTGGAGCAAGAGCACAACAATTGGTCAACGTCCTCTCTGCGT 1017
|||||
Ddb 353 CACAATTATGATCGACTGGAGCAAGAGCACAATAATCTGGTGAACGTTCTCTCTGCGT 412
|||||
QY 1018 GGATATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCG 1077
|||||
Ddb 413 GGACATGTGCTCAACTGGCTGCTGAATGTTTATGATACAGCGGTGCAACGGGACGTATACG 472
|||||
QY 1078 TGTCCTGTCTTTTAAACTGGCATCATTTCCCTGTGTTAAAGCACATTTGGGAAGCAAGTA 1137
|||||
Ddb 473 CGTCTTATCTTTTAAACTGGTGAATTTCCCTGTGTTAAAGCACATTTGGGAAGCAAGTA 532
|||||
QY 1138 CAGATACCTTTTCAAGCAAGTGGCAAGTTTCAACAGGATTTTGTGACCGCAGGCTGGG 1197
|||||
Ddb 533 CAGATACCTTTTCAAGCAAGTGGCAAGTCCACGGGATTTCTGTGACCGGAGACTGGG 592
|||||
QY 1198 CCTCCTCTGTCATGATCTTATCCAAATTTCCAAAGACAGTTGGG 1239
|||||
Ddb 593 CCTGCTT-TACATGATGCAATTCAAATTTCCCGACAGCTGGG 633
|||||

RESULT 13
LOCUS BE201973/c 665 bp mRNA linear EST 30-AUG-2000
DEFINITION fl03h06.x1 zebrafish Research Genetics C32 fin Danio rerio cDNA 3',
similar to TR:Q02295 Q02295 MAJOR DUCHENNE MUSCULAR DYSTROPHY
PROTEIN ; mRNA sequence.
ACCESSION BE201973 GI:8714147
VERSION BE201973.1
KEYWORDS EST.
SOURCE zebrafish.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
1 (bases 1 to 665)
Clark, M., Johnson, S.L., Lehrach, H., Lee, R., Li, F., Marra, M., Eddy
, S., Hillier, L., Kucaba, T., Martin, J., Beck, C., Wylie, T., Underwood
, K., Steptoe, M., Theising, B., Allen, M., Bowers, Y., Person, B.,
Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk, R., Ritter, E.,
Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann, R., Waterston, R.
and Wilson, R.
WashU Zebrafish EST Project 1998
Unpublished (1998)
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@watson.wustl.edu
cDNA Library Preparation: Ning Wu. cDNA Library Arrayed by:
Research Genetics. DNA Sequencing by: Washington University Genome
Sequencing Center Clone distribution: Research Genetics web
address: http://www.researchgenetics.com/
Seq primer: T7 ET from Amersham
High quality sequence stop: 497.
Location/Qualifiers
1..665
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone_lib="Zebrafish Research Genetics C32 fin"
/tissue_type="Fin"
/lab_host="GeneHogs (HS996, a phage-resistant isolate of
DH10B)"
/note="Vector: pT73D-Pac with a modified polylinker;
Site_1: EcoRI; Site_2: NotI; 1st strand cDNA was prepared
from zebrafish(C32) fin, and was then primed with a Not I

```


BASE COUNT	156 a	174 c	176 g	157 t	2 others
ORIGIN	- oligo(dT) primer. Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library is non-normalized. Library was constructed by Ning Wu. NOTE: Clones from this library are only available thru Research Genetics (www.resgen.com). "				

BASE COUNT	156 a	174 c	176 g	157 t	2 others
ORIGIN	Only available thru Research Genetics (www.resgen.com)."				
Query Match	27.1%; Score 358.2; DB 10; Length 665;				
Best Local Similarity	75.9%; Pred. No. 2.7e-96;				
Matches 441; Conservative	0; Mismatches 140; Indels 0; Gaps 0;				

QY	740	TGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAATGTGAGA	799
Db	665	TGGGGACACCCAAAGATGNCAGAACTCTACCAGTCAATTAGCGGATCTCAACAACGTGCGG	606
QY	800	TTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAGGCCCTTTGCTTTGGAT	859
Db	605	TTCTCGGCATACAGSAGCGCAATGAAGCTCAGACGAATGCAGAAAGCCCTCTGTTTGGAT	546
QY	860	CTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGACCAGCAGCAACCTCAAGCAAAATGAC	919
Db	545	CTTCTGAGCATGCCCTGCAGCCTGTGAAGCCTNTGAGCAGCACAATCTCAACACAGAACGAG	486
QY	920	CAGCCCATGGATATCCTGCAGATTATTAATTGTTTGACCACATAATTATGACCGCCTGGAG	979
Db	485	CAGTTTCATGGACATCGTGCAGGTGATCAACTGTCTGACCAGCATCTACGACCGTCTGGAG	426
QY	980	CAAGAGCACAACAATTGGTCAACGTCCTCTCTGCGTGGATATGTGTCTGAACCTGGCTG	1039
Db	425	CAGCAGCACAGCAGCCCTGGTCAACGTGCCCTCTCTGTGTGGACATGTGTCTCAACTGGCTG	366
QY	1040	CTGAATGTTTATGATACGGGACGAACAGGGAGGATCCGTGTCTCTGCTTTTAAAACTGGC	1099
Db	365	CTCAACGTTTACGATACAGGACGAGCTGGGAAGATTCGTACCCCTATCCTTCAAAACACAGGA	306
QY	1100	ATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAGATACCTTTTCAAGCAAGTG	1159
Db	305	ATAATCTCTTTGTGCAAAAGCTCACCTTGAAGATAAGTACAGATTTTATTTTCGAGAGGTG	246
QY	1160	GCAAGTTCAACAGGATTTTGTGACCAGCGCAGGCTGGGCCTCCTTCTGCATGATTCATC	1219
Db	245	GCCAGTGCCACAGGCTTCTGTGACCAGCGCGCCCTCGGCCTCCTCCTGCATGATGCCAAT	186
QY	1220	CAAAATCCAAGACAGTTGGGTGAAGTTGCATCCTTTGGGGGCGAGTAACATTTGAGCCAAAGT	1279
Db	185	CAGATCCCCAGGCAGCTGGGTGAAGTGGCGTCTTTTCGGAGGGAGCAATATTGAGCCCAAGT	126
QY	1280	GTCCGGAGCTGCTTCCAATTTGCTAATAATAAGCCAGAGAT	1320
Db	125	GTGCGCAGCTGCTTTCAGTTCGCCAATAAACAACCCGGAGTT	85

RESULT 14	AL672616	OCUS	DEFINITION
	AL672616	608 bp	linear
	AL672616	XCG-gastrula	silurana tropicalis
		mRNA sequence	cDNA clone TGas054h08 5',
			EST 18-MAR-2002

SOURCE	western clawed frog.
ORGANISM	Silurana tropicalis
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidae; Pipidae; Xenopodinae; Silurana.
REFERENCE	1 (bases 1 to 608)
AUTHORS	Taylor,R., Ashurst,J.L., Croning,M.D.R., Zorn,A.M. and Rogers,J.
TITLE	Sanger Xenopus tropicalis EST project 2002
JOURNAL	Unpublished (2001)
COMMENT	Contact: Taylor R Sanger Centre

Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TGas054h08.p1cSP6
Sequencing primer: p1CSP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
Location/Qualifiers
1. .608
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="TGas054h08"
/clone_lib="XGC-gastrula"
/dev_stage="gastrula (stages 10.5-13 mixed)"
/lab_host="Escherichia coli XLI-blue"
/note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dT primed from 5ug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end"

BASE COUNT	150 a	155 c	151 g	152 t	END.
ORIGIN					
Query Match	25.0%;	Score 330;	DB 9;	Length 608;	
Best Local Similarity	85.0%;	Pred. No. 7.9e-88;			
Matches 369; Conservative	0;	Mismatches 65;	Indels 0;	Gaps 0	
QY 718	CCACGAGACTCAAACAACTT	GTCTGGGACCATCCCAAAAT	GACAGAGCTCTACCAAGTCTTT	777	
Db 175					
QY 778	AGCTGACCTGAATAATGT	CAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	837		
Db 235					
QY 838	GCAGAAGGCCCTTTGCT	TGGATCTCTTGGCCCTGT	CAGCTGCATGTGATGCCCTTGGACCA	897	
Db 295					
QY 898	GCACAACCTCAAGCAAAATG	ACCAGCCCATGGATATCCTGCAGATTATTAATTGTTGAC	957		
Db 355					
QY 958	CACATATTATGACCGCCT	GGAGCAAGAGCACAAATTTGGTCAACGTCCTCTCTGCGT	1017		
Db 415					
QY 1018	GGATATGTGTCTGAAC	TGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCG	1077		
Db 475					
QY 1078	TGTCCTGTCTTTAA	AACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTA	1137		
Db 535					
QY 1138	CAGATACCTTTTCA	1151			
Db 595					
	CAGATACCTTATTCA	608			

```

RESULT 15
BE334408
LOCUS
DEFINITION
    BE334408
    ug97g06.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
    IMAGE:1550362 5' similar to gb:M68859 Mouse dystrophin mRNA, exons
    1-7 and complete cds (MOUSE);, mRNA sequence.
ACCESSION
    BE334408
VERSION
    BE334408.1 GI:9208184
KEYWORDS
    EST.
SOURCE
    house mouse.
    Mus musculus
ORGANISM
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

```


REFERENCE 1 (bases 1 to 532)
AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
JOURNAL Unpublished (1997)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgaps-r@mail.nih.gov
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:955710
Seq primer: -40RP from Gibco
High quality sequence stop: 449.
FEATURES
Location/Qualifiers
1..532
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1550362"
/clone_lib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/note="Vector: p7T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - Oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified p7T3 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT	121 a	151 c	132 g	128 t	
ORIGIN					
Query Match	24.9%	Score 329;	DB 10;	Length 532;	
Best Local Similarity	93.2%;	Pred. No. 1.4e-87;			
Matches 344;	Conservative 0;	Mismatches 25;	Indels 0;	Gaps 0;	
QY 718	CCACGAGACTCAAAACACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAAGTCTTT	777			
Db 164	CCACGAGACCCCAACCACTTGTGGGACCACCCCAAAATGACAGAGCTCTACCAAGTCTTT	223			
QY 778	AGCTGACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	837			
Db 224	AGCTGACCTGAATAATGTCAGGTCTCCCGGTATAGGACTGCCATGAAGCTCAGAAGGCT	283			
QY 838	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTGAGCTGCATGTGATGCCCTTGGACCA	897			
Db 284	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTGAGCTGCATGTGATGCCCTTGGACCA	343			
QY 898	GCACAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAAATGTTTGAC	957			
Db 344	GCACAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAAATGTTTGAC	403			
QY 958	CACATATTATGACCGCCTGGAGCAGACGACACAACAATTTGGTCAACGTCCTCTCTGCGT	1017			
Db 404	TACAATTTATGATCGTCTGGAGCAAGAGCACAACAATCTGGTCAATGTCCCTCTCTGTGT	463			
QY 1018	GGATATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGACAGGAGGATCCG	1077			
Db 464	GGATATGTGTCTGAACCTGGCTTCTCAATGTTTATGATACGGGACGACAGGAGGATCCG	523			
QY 1078	TGTCCTGTC 1086				
Db 524	TGTCCTGTC 532				

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 08:10:12 ; Search time 30.0867 Seconds
(without alignments)
13465.096 Million cell updates/sec

Title: US-09-845-416-10_COPY_1800_3120
Perfect score: 1321
Sequence: 1 cgactttccagcaggttcaga.....ctaataataagccagagatc 1321

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 875442

Minimum DB seq length: 0
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	38.6	2.9	1995	1	US-08-425-069-3
2	38.6	2.9	1995	2	US-08-317-844B-3
3	34.6	2.6	2277	1	US-08-676-967-5
4	34.6	2.6	2277	1	US-08-676-974-5
5	34.6	2.6	2277	2	US-09-098-487-5
6	34.4	2.6	1047	4	US-09-671-950-1
7	34.4	2.6	1047	4	US-09-671-950-3
8	34.4	2.6	1047	4	US-09-671-950-5
9	34.4	2.6	1047	4	US-09-671-950-7
10	34.4	2.6	1047	4	US-09-671-950-9
11	34.4	2.6	1047	4	US-09-671-950-11
12	34.4	2.6	1047	4	US-09-671-950-13
13	34.2	2.6	750	4	US-08-961-527-370
14	34.2	2.6	1620	4	US-08-858-207A-56
15	34.2	2.6	2800	2	US-08-874-138-1
16	34.2	2.6	2800	2	US-08-874-138-5
17	34.2	2.6	2800	4	US-08-879-941-1
18	34.2	2.6	2800	4	US-08-879-941-3
19	34.2	2.6	2800	4	US-09-747-116-1
20	34.2	2.6	2800	4	US-09-747-116-3
21	34	2.6	2235	4	US-09-153-804-2
22	33.8	2.6	1603	1	US-08-625-209A-1
23	33.8	2.6	3489	2	US-08-728-323A-1
24	33.8	2.6	3489	4	US-09-298-568-1
25	33.6	2.5	3527	2	US-08-909-965C-7
26	33	2.5	2169	3	US-08-806-326-5
27	32.8	2.5	2085	2	US-08-668-128B-7

28	32.8	2.5	2085	2	US-08-905-445-7	Sequence 7, Appli
C 29	32.4	2.5	3079	4	US-09-643-597-116	Sequence 116, App
C 30	32.2	2.4	2470	1	US-07-745-206A-14	Sequence 14, Appl
C 31	32.2	2.4	2470	2	US-08-311-363-14	Sequence 14, Appl
32	32	2.4	941	4	US-08-961-527-368	Sequence 368, App
C 33	32	2.4	1493	4	US-09-153-804-7	Sequence 7, Appli
C 34	32	2.4	1875	2	US-09-067-351-4	Sequence 4, Appli
C 35	32	2.4	1875	4	US-09-360-490-4	Sequence 4, Appli
36	31.6	2.4	276	4	US-09-222-575-80	Sequence 80, Appl
37	31.6	2.4	4483	4	US-08-961-527-363	Sequence 363, App
38	31.4	2.4	2610	2	US-09-212-771-1	Sequence 1, Appli
39	31.4	2.4	2610	3	US-09-091-058-1	Sequence 1, Appli
40	31.2	2.4	1512	3	US-08-476-509B-1	Sequence 1, Appli
41	31.2	2.4	1626	3	US-08-348-518C-1	Sequence 1, Appli
42	31	2.3	2790	3	US-08-895-601-2	Sequence 2, Appli
C 43	30.8	2.3	1443	3	US-08-959-381A-3	Sequence 3, Appli
44	30.8	2.3	1599	3	US-08-853-733B-1	Sequence 1, Appli
45	30.8	2.3	1603	3	US-08-675-885-6	Sequence 6, Appli

ALIGNMENTS

RESULT 1
US-08-425-069-3
; Sequence 3, Application US/08425069
; Patent No. 5728810
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Xu, Ming
; APPLICANT: Hinman, Michael B.
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 301 No. 5728810th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/425,069
; FILING DATE: 19-APR-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-106P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 205-8000
; TELEFAX: (703) 205-8050
; TELEX:
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1995 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; IMMEDIATE SOURCE:
; CLONE: p6B
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1785

US-08-425-069-3

Query Match 2.9%; Score 38.6; DB 1; Length 1995;
Best Local Similarity 44.0%; Pred. No. 0.034;
Matches 164; Conservative 0; Mismatches 209; Indels 0; Gaps 0;
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
HYPOTHETICAL: NO
IMMEDIATE SOURCE:
CLONE: p6B
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1785
US-08-317-844B-3

Query Match 2.9%; Score 38.6; DB 2; Length 1995;
Best Local Similarity 44.0%; Pred. No. 0.034;
Matches 164; Conservative 0; Mismatches 209; Indels 0; Gaps 0;
QY 11 GCAGTTTCAGAACGATGTACATAGGGCCCTTCAAGAGGGAATTGAAAACATAAGAA 70
Db 731 GCAGTGCAGCTGCAGACGCCCGCAGCAGGACCTGGACACACAGGACCCCGGAGGATATGGAC 790
QY 71 CCTGTAATCATGACTTCTTGTAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGAA 130
Db 791 CAGGACAACAAAGGACCATCTGACCCGGTAGTCCCGCTGCAGCAGCAGCAGCCCGCAGCAG 850
QY 131 GGACTAGAGAACTCTACGAGAGCCCGCAGAGAGCTGCCTCCTGAGGAGAGAGCCCGCAGAA 190
Db 851 GACCTGGAGGATATGGCCCTGGACAAACAGGACCCCGGAGGATATGGACCCAGGACAACAAG 910
QY 191 GTCACCTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTGAAC 250
Db 911 GACCATCTGGAGCAGGACGTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGGAT 970
QY 251 CTGCACTCCGCTGACTGGCAGAGAGAAATAAGATAGAGACCCCTTGAAGACTCCAGGAACTT 310
Db 971 TAGGAGGTATGGACCAAGGACCAACAAAGTCCAGGAGGATATGGACCCAGGACAACAAGGTC 1030
QY 311 CAAGAGCCACGGATGAGCTGGACCTCAAGCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGATCC 370
Db 1031 CAGGAGGATATGGACCAAGGATAGTGTCTGATCTGACAGCAGCAGCAGCAGCAGCAGCAGCAG 1090
QY 371 TGGCAGCCCGCTGG 383
Db 1091 AAGGACCAGGAGG 1103

RESULT 3
US-08-676-967-5/c
; Sequence 5, Application US/08676967
; Patent No. 5747317
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,967
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055

Query Match 2.9%; Score 38.6; DB 1; Length 1995;
Best Local Similarity 44.0%; Pred. No. 0.034;
Matches 164; Conservative 0; Mismatches 209; Indels 0; Gaps 0;
QY 11 GCAGTTTCAGAACGATGTACATAGGGCCCTTCAAGAGGGAATTGAAAACATAAGAA 70
Db 731 GCAGTGCAGCTGCAGACGCCCGCAGCAGGACCTGGACACACAGGACCCCGGAGGATATGGAC 790
QY 71 CCTGTAATCATGACTTCTTGTAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGAA 130
Db 791 CAGGACAACAAAGGACCATCTGACCCGGTAGTCCCGCTGCAGCAGCAGCAGCCCGCAGCAG 850
QY 131 GGACTAGAGAACTCTACGAGAGCCCGCAGAGAGCTGCCTCCTGAGGAGAGAGCCCGCAGAA 190
Db 851 GACCTGGAGGATATGGCCCTGGACAAACAGGACCCCGGAGGATATGGACCCAGGACAACAAG 910
QY 191 GTCACCTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTGAAC 250
Db 911 GACCATCTGGAGCAGGACGTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGGAT 970
QY 251 CTGCACTCCGCTGACTGGCAGAGAGAAATAAGATGAGACCCCTTGAAGACTCCAGGAACTT 310
Db 971 TAGGAGGTATGGACCAAGGACCAACAAAGTCCAGGAGGATATGGACCCAGGACAACAAGGTC 1030
QY 311 CAAGAGCCACGGATGAGCTGACCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGATCC 370
Db 1031 CAGGAGGATATGGACCAAGGATAGTGTCTGATCTGACAGCAGCAGCAGCAGCAGCAGCAGCAG 1090
QY 371 TGGCAGCCCGCTGG 383
Db 1091 AAGGACCAGGAGG 1103

RESULT 2
US-08-317-844B-3
; Sequence 3, Application US/08317844B
; Patent No. 5989894
; GENERAL INFORMATION:
; APPLICANT: Lewis, Randolph V.
; APPLICANT: Xu, Ming
; APPLICANT: Himman, Michael B.
; TITLE OF INVENTION: ISOLATED DNA CODING FOR SPIDER SILK
; TITLE OF INVENTION: PROTEIN, A REPLICABLE VECTOR AND A TRANSFORMED CELL
; TITLE OF INVENTION: CONTAINING THE ISOLATED DNA, AND PRODUCTS THEREOF
; NUMBER OF SEQUENCES: 62
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Birch, Stewart, Kolasch & Birch
; STREET: 301 No. 5989894th Washington Street
; CITY: Falls Church
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22046
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/317,844B
; FILING DATE: 04-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Murphy Jr., Gerald M
; REGISTRATION NUMBER: 28,977
; REFERENCE/DOCKET NUMBER: 1447-105P
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 241-1300
; TELEFAX: (703) 241-2848
; TELEX: 248345
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1995 base pairs


```
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)343-4341
TELEFAX: (415)343-4342
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 2277 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-676-967-5

Query Match      2.6%; Score 34.6; DB 1; Length 2277;
Best Local Similarity 47.5%; Pred. No. 0.73;
Matches 103; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 308 CTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGA 367
Db 1791 CTGCAGGCTGCGCTGGATGCGCAGCTCCTTCATCTTCAGCTTGCGGCGGTCTCCAGGCT 1732

QY 368 TCCTGGCAGCCCGTGGCGGATCTCCTCATTGACTCTCTCCAAGATCACCTCGAGAAAGTC 427
Db 1731 GAACTCCACGATGGGGCGCTTCAGGGGGCCGAAGATCTCGGGGTGTTGTTGATCAGGCG 1672

QY 428 AAGGCACCTTCGAGGAGAAATTGCGCCTCTGAAAGAGAACGTGAGCCACGTCAATGACCTT 487
Db 1671 CAGGGCCTTCAGGGCGTCTCGTCTCTGGAACCTCGGCGAAGCGTAGCCAGGCTCTG 1612

QY 488 GCTCGCCAGCTTACCACCTTTGGGCATTTCAGCTCTCAC 524
Db 1611 GCCCTTCATGTTGCCGTGCACGCCCTTCAGGTTCGCGC 1575
```

RESULT 4

```
US-08-676-974-5/c
; Sequence 5, Application US/08676974
; Patent No. 5770422
; GENERAL INFORMATION:
; APPLICANT: COLLINS, KATHLEEN
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/676,974
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; S-08-676-974-5
```

```
Query Match      2.6%; Score 34.6; DB 1; Length 2277;
Best Local Similarity 47.5%; Pred. No. 0.73;
Matches 103; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 308 CTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGA 367
Db 1791 CTGCAGGCTGCGCTGGATGCGCAGCTCCTTCATCTTCAGCTTGCGGCGGTCTCCAGGCT 1732

QY 368 TCCTGGCAGCCCGTGGCGGATCTCCTCATTGACTCTCTCCAAGATCACCTCGAGAAAGTC 427
Db 1731 GAACTCCACGATGGGGCGCTTCAGGGGGCCGAAGATCTCGGGGTGTTGTTGATCAGGCG 1672

QY 428 AAGGCACCTTCGAGGAGAAATTGCGCCTCTGAAAGAGAACGTGAGCCACGTCAATGACCTT 487
Db 1671 CAGGGCCTTCAGGGCGTCTCGTCTCTGGAACCTCGGCGAAGCGTAGCCAGGCTCTG 1612

QY 488 GCTCGCCAGCTTACCACCTTTGGGCATTTCAGCTCTCAC 524
Db 1611 GCCCTTCATGTTGCCGTGCACGCCCTTCAGGTTCGCGC 1575
```

RESULT 5

```
US-09-098-487-5/c
; Sequence 5, Application US/09098487
; Patent No. 5917025
; GENERAL INFORMATION:
; APPLICANT: COLLINS, Kathleen
; TITLE OF INVENTION: Human Telomerase
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Science & Technology Law Group
; STREET: 268 Bush Street, Suite 3200
; CITY: San Francisco
; STATE: CA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/098,487
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Osman Ph.D., Richard A
; REGISTRATION NUMBER: 36,627
; REFERENCE/DOCKET NUMBER: UCB96-055
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415)343-4341
; TELEFAX: (415)343-4342
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2277 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-09-098-487-5
```

```
Query Match      2.6%; Score 34.6; DB 2; Length 2277;
Best Local Similarity 47.5%; Pred. No. 0.73;
Matches 103; Conservative 0; Mismatches 114; Indels 0; Gaps 0;

QY 308 CTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGA 367
Db 1791 CTGCAGGCTGCGCTGGATGCGCAGCTCCTTCATCTTCAGCTTGCGGCGGTCTCCAGGCT 1732

QY 368 TCCTGGCAGCCCGTGGCGGATCTCCTCATTGACTCTCTCCAAGATCACCTCGAGAAAGTC 427
Db 1731 GAACTCCACGATGGGGCGCTTCAGGGGGCCGAAGATCTCGGGGTGTTGTTGATCAGGCG 1672
```


RESULT 7
US-09-671-950-3
; Sequence 3, Application US/09671950
; Patent No. 6368830
; GENERAL INFORMATION:
; APPLICANT: Lampe, David
; APPLICANT: Akerley, Brian
; APPLICANT: Rubin, Eric
; APPLICANT: Robertson, Hugh
; TITLE OF INVENTION: Hyperactive Mutants of Himarl Transposase and Methods
; TITLE OF INVENTION: for Using the Same
; FILE REFERENCE: 79-99
; CURRENT APPLICATION NUMBER: US/09/671,950
; CURRENT FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/157,680
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 1047

RESULT 9
US-09-671-950-7
; Sequence 7, Application US/09671950
; Patent No. 6368830
; GENERAL INFORMATION:


```
; APPLICANT: Lampe, David
; APPLICANT: Akerley, Brian
; APPLICANT: Rubin, Eric
; APPLICANT: Robertson, Hugh
; TITLE OF INVENTION: Hyperactive Mutants of Himarl Transposase and Methods
; FILE REFERENCE: 79-99
; CURRENT APPLICATION NUMBER: US/09/671,950
; CURRENT FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/157,680
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 1047
; TYPE: DNA
; ORGANISM: Haematobia irritans
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1044)
US-09-671-950-7
```

```
Query Match          2.6%; Score 34.4; DB 4; Length 1047;
Best Local Similarity 51.3%; Pred. No. 0.52;
Matches 80; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 207 GAAAGCAGGCTGAGGAGGTCAATGAGTGGGAAAAATTGAACCTGCACCTCCGCTGACT 266
Db 50 GAAAAAATACAGTGGAGCAAAACTTGGCTTGATAATGAGTTTCCGGACTCCGCCCCAG 109

QY 267 GGCAGAGAAAAATAGATGAGACCCCTTGAAGACTCCAGGAACCTTCAAGAGGCCACGGATG 326
Db 110 GGAATCAACAATATGATGGTATGCAAAATTCAACCGTGGTGAATGAGCACGGAGG 169

QY 327 AGCTGGACCTCAAGCTGCCCAAGCTGAGGTGATCA 362
Db 170 ACGGTGAACGACGTGGACGCCCGGAAAGGGGTGGTTA 205
```

```
RESULT 10
US-09-671-950-9
; Sequence 9, Application US/09671950
; Patent No. 6368830
; GENERAL INFORMATION:
; APPLICANT: Lampe, David
; APPLICANT: Akerley, Brian
; APPLICANT: Rubin, Eric
; APPLICANT: Robertson, Hugh
; TITLE OF INVENTION: Hyperactive Mutants of Himarl Transposase and Methods
; FILE REFERENCE: 79-99
; CURRENT APPLICATION NUMBER: US/09/671,950
; CURRENT FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/157,680
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 1047
; TYPE: DNA
; ORGANISM: Haematobia irritans
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1044)
US-09-671-950-9
```

```
Query Match          2.6%; Score 34.4; DB 4; Length 1047;
Best Local Similarity 51.3%; Pred. No. 0.52;
Matches 80; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 207 GAAAGCAGGCTGAGGAGGTCAATGAGTGGGAAAAATTGAACCTGCACCTCCGCTGACT 266
Db 50 GAAAAAATACAGTGGAGCAAAACTTGGCTTGATAATGAGTTTCCGGACTCTGCCCCAG 109
```

```
QY 267 GGCAGAGAAAAATAGATGAGACCCCTTGAAGAACTCCAGGAACCTTCAAGAGGCCACGGATG 326
Db 110 GGAATCAACAATATGATGGTATGCAAAATTCAAGCGTGGTGAATGAGCACGGAGG 169

QY 327 AGCTGGACCTCAAGCTGCCCAAGCTGAGGTGATCA 362
Db 170 ACGGTGAACGACGTGGACGCCCGGAAAGAGGTGGTTA 205
```

```
RESULT 11
US-09-671-950-11
; Sequence 11, Application US/09671950
; Patent No. 6368830
; GENERAL INFORMATION:
; APPLICANT: Lampe, David
; APPLICANT: Akerley, Brian
; APPLICANT: Rubin, Eric
; APPLICANT: Robertson, Hugh
; TITLE OF INVENTION: Hyperactive Mutants of Himarl Transposase and Methods
; FILE REFERENCE: 79-99
; CURRENT APPLICATION NUMBER: US/09/671,950
; CURRENT FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/157,680
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 1047
; TYPE: DNA
; ORGANISM: Haematobia irritans
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1044)
US-09-671-950-11
```

```
Query Match          2.6%; Score 34.4; DB 4; Length 1047;
Best Local Similarity 51.3%; Pred. No. 0.52;
Matches 80; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 207 GAAAGCAGGCTGAGGAGGTCAATGAGTGGGAAAAATTGAACCTGCACCTCCGCTGACT 266
Db 50 GAAAAAATACAGTGGAGCAAAACTTGGCTTGATAATGAGTTTCCGGACTCTGCCCCAG 109

QY 267 GGCAGAGAAAAATAGATGAGACCCCTTGAAGAACTCCAGGAACCTTCAAGAGGCCACGGATG 326
Db 110 GGAATCAACAATATGATGGTATGCAAAATTCAAGCGTGGTGAATGAGCACGGAGG 169

QY 327 AGCTGGACCTCAAGCTGCCCAAGCTGAGGTGATCA 362
Db 170 ACGGTGAACGACGTGGACGCCCGGAAAGAGGTGGTTA 205
```

```
RESULT 12
US-09-671-950-13
; Sequence 13, Application US/09671950
; Patent No. 6368830
; GENERAL INFORMATION:
; APPLICANT: Lampe, David
; APPLICANT: Akerley, Brian
; APPLICANT: Rubin, Eric
; APPLICANT: Robertson, Hugh
; TITLE OF INVENTION: Hyperactive Mutants of Himarl Transposase and Methods
; FILE REFERENCE: 79-99
; CURRENT APPLICATION NUMBER: US/09/671,950
; CURRENT FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: 60/157,680
; PRIOR FILING DATE: 1999-10-01
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
```



```

; Sequence 1, Application US/08874138
; Patent No. 5882889
; GENERAL INFORMATION:
; APPLICANT: Wallis, Nicola
; TITLE OF INVENTION: No. 5882889e1 Compounds
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dechert Price & Rhoads
; STREET: 997 Lenox Drive, Building 3, Suite 210
; CITY: Lawrenceville
; STATE: NJ
; COUNTRY: USA
; ZIP: 08543
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/874,138
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Bloom, Allen
; REGISTRATION NUMBER: 29,135
; REFERENCE/DOCKET NUMBER: GM10015
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 609-520-3214
; TELEFAX: 609-520-3259
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2800 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
;
US-08-874-138-1

```

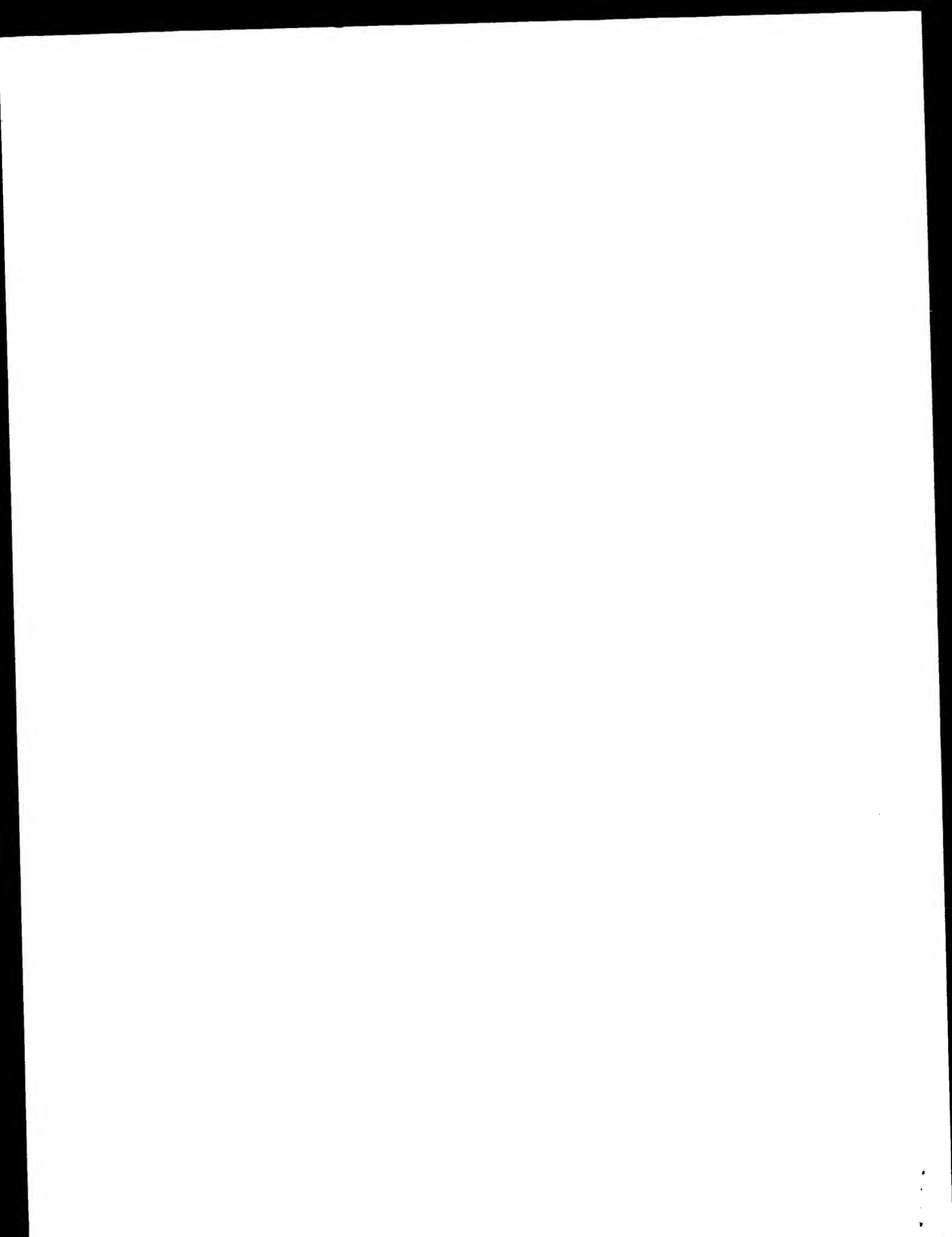
```

Query Match      2.6%; Score 34.2; DB 2; Length 2800;
Best Local Similarity 50.9%; Pred. No. 1.1;
Matches 81; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 976 GGAGCAAGAGCACACAATTGTTGTCACGTCCTCTCTGCGTGGATATGTGCTGAACTG 1035
Db 1685 GGATATGGAGCAGATAAAGAGATCTACGACTCGGTCTTAAGGGAATCCAGTCAGAAAT 1744
QY 1036 GCTGCTGAATGTTTATGATACGGGACGAGGAGGATCCGTGTCCTGCTTTTAAAC 1095
Db 1745 GCAGGACAATAAATATGACCTGGGCAGATTGGTGAATATTCGTGACCGTGCCCTCAAGAG 1804
QY 1096 TGGCATCATTTCCCTGTGTAAGACACATTTGGAAGACAA 1134
Db 1805 TCTCCTAGCTGGAAATTTATAAAGCTAGAGAAAAAGAA 1843

```

Search completed: April 25, 2003, 17:43:52
Job time : 43.0867 secs



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

```
Run on:      April 25, 2003, 17:39:49 ; Search time 87.7709 Seconds
              (without alignments)
              16376.941 Million cell updates/sec
```

```

Title:
Perfect score: 1321
Sequence: 1 CGACTTTCGAGCGATTGAGG
US-09-845-416-10_COPY_1800_3120
updates/sec

```

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 709820 seqs, 544064369 residues

Total number of hits satisfying chosen parameters: 1410746

```
Minimum DB seq length: 0
Maximum DB seq length: 5
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45
```

Database : Published Applications v1.0

```
Published_Applications_NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	54.2	4.1	449	9	US-09-918-995-24084	Sequence 24084, A
2	54.2	4.1	2247	10	US-09-960-253-157	Sequence 157, App
C 3	41.8	3.2	440	9	US-10-184-644-442	Sequence 442, App
C 4	41.8	3.2	440	9	US-10-184-634-442	Sequence 442, App
C 5	38.4	2.9	497	9	US-09-918-995-30349	Sequence 30349, A
6	37.6	2.8	417	9	US-09-918-995-3955	Sequence 3955, Ap
7	37.2	2.8	891	9	US-09-738-626-2757	Sequence 2757, Ap
C 8	37.2	2.8	960	9	US-09-738-626-2758	Sequence 2758, Ap
C 9	36.8	2.8	777	9	US-10-184-644-348	Sequence 348, App
C 10	36.8	2.8	777	9	US-10-184-634-348	Sequence 348, App
11	36	2.7	415	10	US-09-960-352-2580	Sequence 2580, Ap
C 12	35.4	2.7	594	9	US-10-123-155-10	Sequence 10, Appl
13	35.4	2.7	813	9	US-10-184-644-466	Sequence 466, App
14	35.4	2.7	813	9	US-10-184-634-466	Sequence 466, App
15	34.4	2.6	373	10	US-09-867-701-724	Sequence 724, App
16	34.2	2.6	2800	10	US-09-747-116-1	Sequence 1, Appli
17	34.2	2.6	2800	10	US-09-747-116-3	Sequence 3, Appli
18	34	2.6	398	10	US-09-960-352-8059	Sequence 8059, Ap
19	34	2.6	421	10	US-09-960-352-14430	Sequence 14430, A

C	20	34	2.6	448	10	US-09-960-352-5218	Sequence 5218, Ap
C	21	34	2.6	2529	10	US-09-969-708-575	Sequence 575, App
C	22	33.6	2.5	4090	10	US-09-925-301-536	Sequence 536, App
	23	33.2	2.5	3489	10	US-09-815-242-7731	Sequence 7731, Ap
	24	33	2.5	4270	9	US-09-954-531-1352	Sequence 1352, Ap
	25	33	2.5	4270	10	US-09-880-107-2427	Sequence 2427, Ap
	26	32.6	2.5	404	10	US-09-960-352-10622	Sequence 10622, A
C	27	32.6	2.5	734	9	US-10-184-644-458	Sequence 458, App
C	28	32.6	2.5	734	9	US-10-184-634-458	Sequence 458, App
C	29	32.4	2.5	470	9	US-09-918-995-24845	Sequence 24845, A
C	30	32.4	2.5	1455	9	US-09-764-868-348	Sequence 348, App
C	31	32.4	2.5	3079	10	US-09-735-705-116	Sequence 116, App
C	32	32.4	2.5	3079	10	US-09-850-716A-116	Sequence 116, App
C	33	32.4	2.5	3079	10	US-09-897-778-116	Sequence 116, App
C	34	32.4	2.5	3544	9	US-09-974-298-34	Sequence 34, Appl
C	35	32.4	2.5	3544	10	US-09-919-172-12	Sequence 12, Appl
	36	32.2	2.4	242	9	US-09-796-692-9375	Sequence 9375, Ap
	37	32.2	2.4	355	10	US-09-960-352-2696	Sequence 2696, Ap
C	38	32.2	2.4	498	9	US-09-918-995-21329	Sequence 21329, A
C	39	32	2.4	463	9	US-09-918-995-17920	Sequence 17920, A
C	40	32	2.4	1365	9	US-09-974-298-29	Sequence 29, Appl
C	41	32	2.4	1365	12	US-10-044-090-70	Sequence 70, Appl
C	42	32	2.4	1493	10	US-09-964-824A-549	Sequence 549, App
C	43	32	2.4	1493	10	US-09-967-768A-143	Sequence 143, App
C	44	32	2.4	1499	9	US-10-102-806-45	Sequence 45, Appl
C	45	32	2.4	1517	10	US-09-964-824A-548	Sequence 548, App

ALIGNMENTS

```

RESULT 1
US-09-918-995-24084
; Sequence 24084, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756 FROM VARIOUS CDNA LIBRARIES
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 24084
; LENGTH: 449
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(449)
; OTHER INFORMATION: n = A,T,C or G
US-09-918-995-24084

```

Query Match	4.1%;	Score 54.2;	DB 9;	Length 449;
Best Local Similarity	51.9%;	Pred. No. 7.2e-08;		
Matches 122.	Concentration			

		122	Conservative	0;	Mismatches	113;	Indels	0;	Gaps						
QY	1028	CTGA	ACTG	GCTG	CTGA	ATGTT	TATG	ATAC	GGGAC	GAAC	AGGAGG	ATCC	GTGTC	TGTCT	1087
Db	198	CTCA	ACTT	TATG	ATGCT	GCAT	ATG	ACAG	TGAG	GGCG	GGCA	AGTTG	ACGG	TATTTC	257
QY	1088	TTTAA	AACTG	GCAT	CA	TTCC	CTGT	GTA	AGC	ACAT	TTGGA	AGACA	AGTAC	AGATAC	CTT 1147
Db	258	GT	TAA	GCT	ATGTT	AGCA	ACC	ATG	TGG	TGG	AAAA	ATGCT	GGACA	AAATG	AGATATGTT 317
QY	1148	TTCA	AGCA	AGTGG	CAAG	TTCA	ACAG	GATTT	TGTG	ACCA	GCGC	AGGCTGGG	CCCTC	CTTCTG 1207	
Db	318	TTCT	CCC	CAG	ATG	TCC	CAATG	CGCTT	AAATG	ATAT	TAG	CAAG	TTTG	ACCAGTTTCTG 377	

1208 C A G A T T C T A T C C A A A T T C C A A G A C A G T T G G G T G A A G T T G C A T C C T T T G G G G C A 1262

ORGANISM: Homo Sapien

608 CATGAAGCCACAGGGACTTT

608 CATGAAGCCACAGGGACTTT

Db 369 T...CYKHK...NHRCDWCY.WC.YYCASNKW...K.YN.A...WC.HTB.A...SB.MCMMR 310
QY 668 GGTCCCTGGGAGAGAGCCATCTCGCCCAACAAAGTGCCTACTATATCAACACGAGACT 727
Db 309 R.YS..T.M.RBKD.BAS.MNR..SDMC.T.W.SBB.CYHMB...HN.S.RMN...BMH 250
QY 728 CAAACAACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCACTCTTTAGCTGACCTG 787
Db 249 KW..HAKCDHS...BN.YH..CY...WMAKRM.BAHBTM.THY...ST.CRDDC.BH.Y. 190
QY 788 AATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAGGCC 847
Db 189 DR.TT.HTYCSM.SH.THSBNT.N.HA.BNMR..GRA.YBC...B.GC.WRRKRBMYMT 130
QY 848 CTTTGTGCTGATCTCTTGAGCCTGTGAGCTGCATGTGATGCCTTGGACACGACACACCTC 907
Db 129 MA.ABW...SCNWW..K.SHMNBC.H.SMNAA.HWKRCN.K.DB.GDYCAK.ARRN.RS. 70
QY 908 AAGCAAAATGACCAGGCCA 926
Db 69 S..H.A.RHBMBCYSB...M 51

RESULT 5
US-09-918-995-30349/c
; Sequence 30349, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 30349
; LENGTH: 497
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(497)
; OTHER INFORMATION: n = A,T,C or G
JS-09-918-995-30349

Query Match 2.9%; Score 38.4; DB 9; Length 497;
Best Local Similarity 55.1%; Pred. No. 0.021;
Matches 75; Conservative 0; Mismatches 61; Indels 0; Gaps 0;
QY 249 ACCTGCACCTCGCTGACTGGCAGAGAGAAAATAGATGAGACCTTGAAAGACTCCAGGAAC 308
Db 408 ACCTGCATCGCGGGGGGATGGACCATCAGGGGAAAAGACCTGGAGCGGCTCCCCCAGA 349
QY 309 TTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGAT 368
Db 348 GTGCAGTGGCCAAAGACCTCTGGAGTCCCTCAGCACCAAGCAGGACGACTCACAGCCT 289
QY 369 CCTGGCAGCCCGTGGG 384
Db 288 CTGGCCAGGTGCTGGG 273

RESULT 6
S-09-918-995-3955
; Sequence 3955, Application US/09918995
; Publication No. US20030073623A1
; GENERAL INFORMATION:
; APPLICANT: Hyseq, Inc.
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED

; TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
; FILE REFERENCE: 20411-756
; CURRENT APPLICATION NUMBER: US/09/918,995
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US/09/235,076
; PRIOR FILING DATE: 1999-01-20
; NUMBER OF SEQ ID NOS: 38054
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3955
; LENGTH: 417
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-918-995-3955

Query Match 2.8%; Score 37.6; DB 9; Length 417;
Best Local Similarity 53.1%; Pred. No. 0.035;
Matches 104; Conservative 0; Mismatches 89; Indels 3; Gaps 1;
QY 169 TCCTGAGGAGAGAGCCAGCAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAA 228
Db 154 TGCTGAGGAGAGAGCCAGAGGCCATCAGCGGATGCCCGCATGTGCGAGAGGAGTCAA 213
QY 229 TACTGAGTGGGAAAAATGAAACCTGCACTCCGCTGACTGGCAGAGAAAAATAGATGAGAC 288
Db 214 GAAGGAGCAGGACA---CCAGCGCCACCTGGAGCGCATGAAGAAGAACATGGAACAGAC 270
QY 289 CCTTGAAGAGACTCCAGGAACCTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCA 348
Db 271 CATTAAGGACCTGCAGCACCGGCTGGACGAGCCGAGGATCGCCCTCAAGGGCGGCAA 330
QY 349 AGCTGAGGTGATCAAG 364
Db 331 GAAGCAGCTGCAGAAAG 346

RESULT 7
US-09-738-626-2757
; Sequence 2757, Application US/09738626
; Publication No. US20020197605A1
; GENERAL INFORMATION:
; APPLICANT: NAKAGAWA, SATOSHI
; APPLICANT: MIZOGUCHI, HIROSHI
; APPLICANT: ANDO, SEIKO
; APPLICANT: HAYASHI, MIKIRO
; APPLICANT: OCHIAI, KEIKO
; APPLICANT: YOKOI, HARUHIKO
; APPLICANT: TATEISHI, NAKO
; APPLICANT: SENOH, AKIHIRO
; APPLICANT: IKEDA, MASATO
; APPLICANT: OZAKI, AKIO
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-125
; CURRENT APPLICATION NUMBER: US/09/738,626
; CURRENT FILING DATE: 2000-12-18
; PRIOR APPLICATION NUMBER: JP 99/377484
; PRIOR FILING DATE: 1999-12-16
; PRIOR APPLICATION NUMBER: JP 00/159162
; PRIOR FILING DATE: 2000-04-07
; PRIOR APPLICATION NUMBER: JP 00/280988
; PRIOR FILING DATE: 2000-08-03
; NUMBER OF SEQ ID NOS: 7059
; SOFTWARE: PatentIn ver. 3.0
; SEQ ID NO 2757
; LENGTH: 891
; TYPE: DNA
; ORGANISM: Corynebacterium glutamicum
US-09-738-626-2757

Query Match 2.8%; Score 37.2; DB 9; Length 891;
Best Local Similarity 48.9%; Pred. No. 0.078;
Matches 130; Conservative 0; Mismatches 133; Indels 3; Gaps 1;
QY 1054 TACGGACGAACAGGGAGGATCCGTGCTCTTTAAAACTGGCATCATTTCCCTGTG 1113

RESULT 11

```

; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; TITLE OF INVENTION: ACIDS ENCODING THE SAME
; FILE REFERENCE: P3330RIC30
; CURRENT APPLICATION NUMBER: US/10/123,155
; CURRENT FILING DATE: 2002-04-15
; Prior Application removed - See Palm or File Wrapper
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 10
; LENGTH: 594

```


Db 3 WRPRRARGTPLL LLLLLLLLWVPVGAGVLQGHIPGPVTPHWVLDGQPWRTVSLEEPVSK 62
QY 121 GCCTTTGGAAGGACTAGAGAAACTCTACAGGAGCCAGAGAGCTGCCTCCTGAGGAGAG 180
Ob 63 PDMGLVALEAEGQELLLELEKHNRLAPGYIETHYGPDPQPVVLAAPNHTDCHYQGRVRG 122
QY 181 AGCCAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGA 240
Db 123 FPDWVVLCTCSGMSGLITLSRNASYLRLPWPWRGSKDFSTHEIFRMEQLLTWKGTGCHR 182
QY 241 AAAATGAACCTGCACCTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAAGACT 300
Db 183 DPGNKAGMTSLPGGPQSRGRREARRTRKYLELYIVADHTLFLTRHNLNHTKORLLEVAN 242
QY 301 CCAGGAACCTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGAT 360
Db 243 YVDQLRLTDIQVALTGLEVWTERDRSRVTQDANATLWAFLOWRRGLWAORPHDSAQLLT 302
QY 361 CAAGGGATCCTGGCAGCCCGTGGCGGATCTCCTCACTGACTCTCTCCAAGATCACCTCGA 420
Db 303 GRAFGATVGLAPVEGMCRAESSGGVSTDHSELPIGAAATMAHEIGHSLGLSHDPDGCVC 362
QY 421 GAAAGTCAAGGCACCTTCGAGGAGAAATTCGCCCTCTGAAAGAGAACGTGAGCCACGTCAA 480
Db 363 EAAESGGCVMAAATGHPFPRVFSACSRRLRAFRKGGGACLSNAPDPGLPVPALCGN 422
QY 481 TGACCTTGCTGCCAGCTTACCACCTTTGGGCAATTCAGCTCTCACCGTATACCTCAGCAC 540
Db 423 GFVEAGEECDCGPGQECRDLCFAHNCSLRPGAQCAHGDCCVRCLLKPAGALCRQAMGDC 482
QY 541 TCTGGAAGACCTGAACACACAGATGGAAGCTTCTGAGGTGGCCGCTCGAGGACCGAGTCAG 600
Db 483 DLPEFCTGTSSHCPDPDVYLLDGSPCARGSGYCWDCACPTLEQQCCQQLWGPESHPEACF 542
QY 601 GCAGCTGCATGAAG 614
Db 543 QVNSAGDAHNGC 556

RESULT 15

JS-09-867-701-724

Sequence 724, Application US/09867701

Patent No. US20020132237A1

GENERAL INFORMATION:

APPLICANT: Aglate, Paul A.

APPLICANT: Jones, Robert

APPLICANT: Harlocker, Susan L.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER

FILE REFERENCE: 210121.497

CURRENT APPLICATION NUMBER: US/09/867,701

CURRENT FILING DATE: 2001-05-29

NUMBER OF SEQ ID NOS: 10912

SOFTWARE: FastSEQ for Windows Version 4.0

SEQ ID NO 724

LENGTH: 373

TYPE: DNA

ORGANISM: Homo sapien

FEATURE:

NAME/KEY: misc_feature

LOCATION: (1)...(373)

OTHER INFORMATION: n = A,T,C or G

JS-09-867-701-724

Query Match

Best Local Similarity 2.6%; Score 34.4; DB 10; Length 373;

Matches 54.8%; Pred. No. 0.41;

Conservative 0; Mismatches 56; Indels 0; Gaps 0;

Y 1122 ATTTGGAAGACAAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTG 1181

Y 64 ACTAGTAACCTCAAGAACAGATGTAGATTAAACAAGTGACAAGTGACGCAAAACTCTTTA 123

Y 1182 ACCAGCGCAGGCTGGGCCTCCTTCTGATGATCTATCCAAATTTCAAGACAGTTGGGTG 1241

Db 124 TTAGGATTTTAATGATCTTAATTTTGCAATGATCTGTCAATAATTCAACAAGAAAGCTTA 183
QY 1242 AAGT 1245
Db 184 AAGT 187
Search completed: April 25, 2003, 23:58:21
Job time : 94.9375 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 07:56:07 ; Search time 3479.7 seconds
(without alignments)
17070.103 Million cell updates/sec

Title: US-09-845-416-12_COPY_960_3000
Perfect score: 2041
Sequence: 1 tcttcacagcatttggaag.....ttctgcatgattctatccaa 2041

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 3813070

Minimum DB seq length: 0
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

- 1: gb_ba:*
- 2: gb_htg:*
- 3: gb_in:*
- 4: gb_om:*
- 5: gb_ov:*
- 6: gb_pat:*
- 7: gb_ph:*
- 8: gb_pl:*
- 9: gb_pr:*
- 10: gb_ro:*
- 11: gb_sts:*
- 12: gb_sy:*
- 13: gb_un:*
- 14: gb_vi:*
- 15: em_ba:*
- 16: em_fun:*
- 17: em_hum:*
- 18: em_in:*
- 19: em_mu:*
- 20: em_om:*
- 21: em_or:*
- 22: em_ov:*
- 23: em_pat:*
- 24: em_ph:*
- 25: em_pl:*
- 26: em_ro:*
- 27: em_sts:*
- 28: em_un:*
- 29: em_vi:*
- 30: em_htg_hum:*
- 31: em_htg_inv:*
- 32: em_htg_other:*
- 33: em_htg_mus:*
- 34: em_htg_pln:*
- 35: em_htg_rod:*
- 36: em_htg_nam:*
- 37: em_htg_vrt:*
- 38: em_sy:*
- 39: em_htgo_hum:*
- 40: em_htgo_mus:*
- 41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1257	61.6	4402	6	E30219	E30219 Shortened d
2	1136.6	55.7	4402	6	E30220	E30220 Shortened d
3	715.6	35.1	2654	5	FSCDYSTRO	M37645 Torpedo cal
4	671.4	32.9	4075	6	E30221	E30221 Shortened d
5	667.2	32.7	3747	6	E30218	E30218 Shortened d
6	613.6	30.1	3163	6	E30223	E30223 Shortened d
7	598.2	29.3	3521	5	AF339031	AF339031 Danio rer
8	585.8	28.7	3275	10	MUSDYS	M18025 Mouse dystro
9	554	27.1	3161	10	MMGUTRPH	X83506 M.musculus
10	528.4	25.9	3073	10	AF195788	AF195788 Rattus no
11	528.4	25.9	3140	10	AF195787	AF195787 Rattus no
12	527.4	25.8	3499	9	HSU43519	U43519 Human dystro
13	505	24.7	2110	9	HUMDMDX	M92650 Human Duch
14	505	24.7	4658	9	BC028720	BC028720 Homo sapi
15	468	22.9	630	9	HSDMDf1	X06179 Human fetal
16	460.4	22.6	1240	10	RNAPDY3	X69767 R.norvegicu
17	394.8	19.3	3172	6	E30222	E30222 Shortened d
18	348.4	17.1	3650	3	SPAJ3356	AJ223356 Strongylo
19	339.6	16.6	1961	5	XLDYSTROP	X99700 X.laevis mR
20	318.8	15.6	1195	5	AF339032	AF339032 Danio rer
21	304.8	14.9	1993	5	SCDYSTROP	X99702 S.caniculua
22	273.4	13.4	500	9	HUMDYSTR15	L05649 Homo sapien
23	265.8	13.0	1898	10	BC024140	BC024140 Mus muscu
24	262.2	12.8	506	10	AB011666	AB011666 Rattus no
25	241.6	11.8	384	10	RNADY1	X65468 R.norvegicu
26	229.6	11.2	1428	5	SCU43517	U43517 Scyliorhinu
27	222	10.9	1482	5	XLUTROPHI	X99701 X.laevis mR
28	216	10.6	1737	5	SCUTROPHI	X99703 S.caniculua
29	213.4	10.5	1966	9	HSDMDA1	X06178 Human adult
30	203	9.9	351	9	HUMDYSTR06	M86889 H.sapiens d
31	198.2	9.7	1500	6	AX107969	AX107969 Sequence
32	198.2	9.7	1500	6	AX107970	AX107970 Sequence
33	194.4	9.5	1440	5	GSU43518	U43518 Gobius sp.
34	193.2	9.5	645	9	AF213410	AF213410 Homo sapi
35	187	9.2	525	9	AF213411	AF213411 Homo sapi
36	186.8	9.2	1428	10	MMU43520	U43520 Mus musculu
37	179.8	8.8	374	10	S62620	S62620 (mdx3Cv)-dy
38	176.2	8.6	1431	3	PSDYSTROP	X99738 Pectinidae
39	174.4	8.5	1338	3	BLDYSTROP	X99736 B.lanceolat
40	173.8	8.5	1413	3	ASDYSTROP	X99737 Asteroidea
41	157	7.7	300	9	S38776	S38776 Homo sapien
42	153.8	7.5	1729	3	CIDYSTRO	X99756 C.intestina
43	152.8	7.5	352	9	AF213412	AF213412 Homo sapi
44	151	7.4	332	9	HUMDYST01	M86884 H.sapiens d
45	149.4	7.3	989	9	HSDYSTX60	Z11860 H.sapiens g

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL
E30219	E30219	Shortened dystrophin.	E30219	E30219.1	GI:13017026	JP 1999318467-A/2.	unidentified.	unclassified.	1 (bases 1 to 4402)	Sinichi,T.	Shortened dystrophin
											Patent: JP 1999318467-A 2 24-NOV-1999;
											SCIENCE & TECH AGENCY,NATIONAL CENTER OF NEUROLOGY AND PSYCHIATRY

COMMENT OS Unidentified
PN JP 1999318467-A/2
PD 24-NOV-1999
PF 08-MAY-1998 JP 1998142134
PR
PI SINICHI TAKEDA
PC C12N15/09,A61K48/00,C12N15/00
CC Strandedness: Both;
CC Topology: Linear;
PH Key Location/Qualifiers
FT source 1..4402
/organism='Unidentified'.
FEATURES
source Location/Qualifiers
1..4402
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 1329 a 1000 c 1019 g 1054 t
ORIGIN
Query Match 61.6%; Score 1257; DB 6; Length 4402;
Best Local Similarity 81.0%; Pred. No. 2.9e-307;
Matches 1654; Conservative 0; Mismatches 0; Indels 387; Gaps 1;
QY 1 TCCTTCACAGCATTTGGAAGCTCCTGAAGACACAGTCAATTTGGCAGTTTCATTGATGGAGAG 60
DB 1155 TCCTTCACAGCATTTGGAAGCTCCTGAAGACACAGTCAATTTGGCAGTTTCATTGATGGAGAG 1214
QY 61 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAGTATTATCGTGGCTTCTTTC 120
DB 1215 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAGTATTATCGTGGCTTCTTTC 1274
QY 121 TGCTGAGGACACATTGCAAGCACAGGAGAGATTTCATATGATGTGGAAGTGGTGAAGA 180
DB 1275 TGCTGAGGACACATTGCAAGCACAGGAGAGATTTCATATGATGTGGAAGTGGTGAAGA 1334
QY 181 CCAGTTTCATCTATGAGGGGTACATGATGGATTTGACAGCCCATCAGGGCCGGTTGG 240
DB 1335 CCAGTTTCATCTATGAGGGGTACATGATGGATTTGACAGCCCATCAGGGCCGGTTGG 1394
QY 241 TAATATTCTACAATTGGGAAGTAAGCTGATTGGAACAGGAAATATATCAGAAGATGAAGA 300
DB 1395 TAATATTCTACAATTGGGAAGTAAGCTGATTGGAACAGGAAATATATCAGAAGATGAAGA 1454
QY 301 AACTGAAGTACAAAGAGCAGATGAATCTCCTAAATTCAAGATGGGAATGCCTCAGGGTAGC 360
DB 1455 AACTGAAGTACAAAGAGCAGATGAATCTCCTAAATTCAAGATGGGAATGCCTCAGGGTAGC 1514
QY 361 TAGCATGGAACAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT 420
DB 1515 TAGCATGGAACAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT 1574
QY 421 GAAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGAAAGAAACAGGAAATGGAGGAAGA 480
DB 1575 GAAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGAAAGAAACAGGAAATGGAGGAAGA 1634
QY 481 GCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCAAGTACAAACAAATTAAGGTGCTTCA 540
DB 1635 GCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCAAGTACAAACAAATTAAGGTGCTTCA 1694
QY 541 AGAAGATCTAGAACAAAGAACAGTCAAGTCAATTTCTCTCAGTCAATGTTGGTGGTAGT 600
DB 1695 AGAAGATCTAGAACAAAGAACAGTCAAGTCAATTTCTCTCAGTCAATGTTGGTGGTAGT 1754
QY 601 TGATGAATCTAGTGGAGATCAGGCAACTGCTGCTTTTGAAGAACAACCTTAAGSTATTGGG 660
DB 1755 TGATGAATCTAGTGGAGATCAGGCAACTGCTGCTTTTGAAGAACAACCTTAAGSTATTGGG 1804
QY 661 AGATCGATGGGCAACACATCTGTAGATGGACAGAACCCGCTGGGTCTTTTACAAAGACAG 720
DB 1805 ----- 1804
QY 721 TTCTGACCAGTGGGAAGCGTCTGCACCTTTCTCTGCAGGAACTTCTGGTGTGGCTACAGCT 780

Db 1805 ----- 1804
QY 781 GAAAGATGATGAATTAAGCGCGCAGGCACCTATTGGAGGGGACTTTCCAGCAGTTTCAGAA 840
Db 1805 ----- 1804
QY 841 GCAGAACGATGTACATAGGGCCTTCAAGAGGGAATTGAAAACTTAAAGAACCTGTATCAT 900
Db 1805 ----- 1804
QY 901 GAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAGGAGTAGAGAA 960
Db 1805 ----- 1804
QY 961 ACTCTACCAGGAGCCAGAGAGCTGCCTCTCTGAGGAGAGAGAGAGAGAGAGAGAGAGAG 1020
Db 1805 ----- 1804
QY 1021 TCTACGAAAGCAGGCTGAGGAGGTCATTAATGAGTGGGAAAAATTTGAACCTGCACTCCGC 1080
Db 1805 -----AGGAGGTCATTAATGAGTGGGAAAAATTTGAACCTGCACTCCGC 1847
QY 1081 TGACTGGCAGAGAAAAATAGATGAGAGACCTTTGAAAGACTCCAGGAACCTTCAAGAGGCCAC 1140
Db 1848 TGACTGGCAGAGAAAAATAGATGAGAGACCTTTGAAAGACTCCAGGAACCTTCAAGAGGCCAC 1907
QY 1141 GGATGAGCTGGACCTCAAGCTGGCCCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGT 1200
Db 1908 GGATGAGCTGGACCTCAAGCTGGCCCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGT 1967
QY 1201 GGGCGATCTCCTCAATTAATGACTCTCTCCAAGATCACCTCGAGAGAAAGTCAAGGCACCTTCGAGG 1260
Db 1968 GGGCGATCTCCTCAATTAATGACTCTCTCCAAGATCACCTCGAGAGAAAGTCAAGGCACCTTCGAGG 2027
QY 1261 AGAAATTGGCCCTCTGAAAGAGAAACGTGAGCCACGTCAATGACCTTGTCTGCCAGCTTAC 1320
Db 2028 AGAAATTGGCCCTCTGAAAGAGAAACGTGAGCCACGTCAATGACCTTGTCTGCCAGCTTAC 2087
QY 1321 CACTTTGGGCATTCAGCTCTCACCGTATAAACCTCAGCACTCTGGAAGACCTGAACACCAG 1380
Db 2088 CACTTTGGGCATTCAGCTCTCACCGTATAAACCTCAGCACTCTGGAAGACCTGAACACCAG 2147
QY 1381 ATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGGAGTCAGGAGTCATGAAGCCACAG 1440
Db 2148 ATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGGAGTCATGAAGCCACAG 2207
QY 1441 GGACTTTGGTCCAGCATCTCAGCACTTCTTTCCACGTCTGTCCAGGGTCCCTGGGAGAG 1500
Db 2208 GGACTTTGGTCCAGCATCTCAGCACTTCTTTCCACGTCTGTCCAGGGTCCCTGGGAGAG 2267
QY 1501 AGCCATCTGCCAAACAAAGTGCCTTACTATATCAACCCAGAGACTCAAAACAACTTGTCTG 1560
Db 2268 AGCCATCTGCCAAACAAAGTGCCTTACTATATCAACCCAGAGACTCAAAACAACTTGTCTG 2327
QY 1561 GGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATTAATGTGATTT 1620
Db 2328 GGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATTAATGTGATTT 2387
QY 1621 CTCAGCTTATAGGACTGCCATGAACCTCCGAAAGACTGCAGAAAGGCCCTTTGCTTGGATCT 1680
Db 2388 CTCAGCTTATAGGACTGCCATGAACCTCCGAAAGACTGCAGAAAGGCCCTTTGCTTGGATCT 2447
QY 1681 CTTGAGCCTGTGAGTGCATGTGATGCCCTTGGACCAAGCAACCTCAAGCAAAATGACCA 1740
Db 2448 CTTGAGCCTGTGAGTGCATGTGATGCCCTTGGACCAAGCAACCTCAAGCAAAATGACCA 2507
QY 1741 GCCCATGGATATCTGCAGATTTAATTTGTTGACCACTATTTATGACCCGCTTGGAGCA 1800
Db 2508 GCCCATGGATATCTGCAGATTTAATTTGTTGACCACTATTTATGACCCGCTTGGAGCA 2567
QY 1801 AGAGCACACAACATTTGGTCAACGTCCCTCTCTGCGTGGATATGTGTCTGAACCTGCTGCT 1860
Db 2568 AGAGCACACAACATTTGGTCAACGTCCCTCTCTGCGTGGATATGTGTCTGAACCTGCTGCT 2627

2088 CACTTTGGGCATTGAGCTCTCACCGTATTAACCTCAGCACTCTGGAAGACCTGAACACACAG 2147

1381 ATGGAAGCTTCTGCAGGTGGCGTCGAGGACCGAGTCAAGCAGCTGCATGAAGCCACAG 1440

2148 ATGGAAGCTTCTGCAGGTGGCGTCGAGGACCGAGTCAAGCAGCTGCATGAAGCCACAG 2207

1441 GGACTTTGGTCCAGCATCTCAGCACTTTCTTCCACGCTGTGTCCAGGGTCCCTGGGAGAG 1500

2208 GGACTTTGGTCCAGCATCTCAGCACTTTCTTCCACGCTGTGTCCAGGGTCCCTGGGAGAG 2267

1501 AGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACACGAGACTCAAAACAACTTGCTG 1560

2268 AGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACACGAGACTCAAAACAACTTGCTG 2327

1561 GGACCATCCCAAAATGACAGAGCTCTACCAAGTCTTTAGCTGACCTGAATAATGTGAGAT 1620

2328 GGACCATCCCAAAATGACAGAGCTCTACCAAGTCTTTAGCTGACCTGAATAATGTGAGAT 2387

1621 CTCAGCTTATAGGACTGCCATGAAACTCCGAAAGACTGCAGAGGCCCTTTGCTTGGATCT 1680

2388 CTCAGCTTATAGGACTGCCATGAAACTCCGAAAGACTGCAGAGGCCCTTTGCTTGGATCT 2447

1681 CTTGAGCCTGTGAGCTGCATGTGATGCCTTGACCAAGCAACCTCAAGCAAAATGACCA 1740

2448 CTTGAGCCTGTGAGCTGCATGTGATGCCTTGACCAAGCAACCTCAAGCAAAATGACCA 2507

1741 GCCCATGGATATCCTGCAGATTATTAATTGTTTGACCACTATTATGACCGCCTGGAGCA 1800

2508 GCCCATGGATATCCTGCAGATTATTAATTGTTTGACCACTATTATGACCGCCTGGAGCA 2567

1801 AGAGCAACAATAATTGGTCAACGTCCTCTGCGTGGATATGTGTGAACTGGCTGCT 1860

2568 AGAGCAACAATAATTGGTCAACGTCCTCTGCGTGGATATGTGTGAACTGGCTGCT 2627

1861 GAATGTTTATGATACGGGACGAACAGGAGGATCCGTGCTCTTTTAAACTGGCAT 1920

2628 GAATGTTTATGATACGGGACGAACAGGAGGATCCGTGCTCTTTTAAACTGGCAT 2687

1921 CATTTCCCTGTGTAAGACACATTTGGAAGACAAGTACAGATACCTTTTCAAGCAAGTGGC 1980

2688 CATTTCCCTGTGTAAGACACATTTGGAAGACAAGTACAGATACCTTTTCAAGCAAGTGGC 2747

1981 AAGTTCAACAGGATTTGTGACCAAGCGAGGCTGGGCTCCTCTGCTGATGATTTATCCA 2040

2748 AAGTTCAACAGGATTTGTGACCAAGCGAGGCTGGGCTCCTCTGCTGATGATTTATCCA 2807

2041 A 2041

2808 A 2808

RESULT 3

FSCDYSTRO 2654 bp mRNA linear VRT 08-SEP-1993

LOCUS Torpedo californica dystrophin mRNA, 3' end.

DEFINITION M37645

ACCESSION M37645.1 GI:397970

VERSION 1

KEYWORDS Torpedo californica, cdna to mRNA.

SOURCE Torpedo californica

ORGANISM Torpedo californica

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes; Elasmobranchii; Squalea; Hypnosqualea; Pristiogaster; Batoidae; Torpediniformes; Torpedinidae; Torpedinidae; Torpedo.

1 (bases 1 to 2654)

Yeadon, J.E., Lin, H., Dyer, S.M., and Burden, S.J.

Dystrophin is a component of the subsynaptic membrane

J. Cell Biol. 115 (4), 1069-1076 (1991)

92064638

1720119

2 (bases 1 to 2654)

Ravin, A.J., Dyer, S.M., Yeadon, J.E. and Burden, S.J.

Multiple dystrophin isoforms are associated with the postsynaptic membrane of Torpedo electric organ

J. Physiol. (Paris) 85 (3), 131-133 (1991)

92291902

1818109

On Sep 8, 1993 this sequence version replaced gi:213232.

Draft entry and computer-readable sequence for [Unpublished (1990)] kindly submitted

by S.J.Burden, 06-AUG-1990.

Biology Dept, 16-820

MIT

Cambridge, MA 02139.

Location/Qualifiers

1. .2654

/organism="Torpedo californica"

/db_xref="taxon:7787"

1. .2654

/gene="dystrophin"

<1. .2654

/gene="dystrophin"

/codon_start=3

/product="dystrophin"

/protein_id="AAC38002.1"

/db_xref="GI:397971"

/translation="SGEQWKRQLQISLDLFTWMLNKLNDLRRQPIGGDAPTVCCQNDVHRIFKRELKAKEPVVMSALDTVHLFLADPAIRAPESLLTGPKEIKPEENIQNVAKRI RKYAEVKEVWDKLSNRSDWKRIEALKRLLEQDLMDELNLKLRQAEALIKDTWQP VGDLIDSLODHIKVKVFAEIAKPMKENVTHMDLASQFTPPDIQLSPVNLNLEDL NTRWKLLOVSDIELLQKHEHRDFTGPTSHFLSTSVQGPWERAI SPKVPYIINHQT QTTCDWHPKMTELYQSLADLNVRFSAYRTAMKLRRLQKALCLDLSLPSACEAFDQH NLKQNDQLLDILEI INCLTSIYDRLEQHSNLSVPLCVDMCLNLLNVYDTGRTGKI RVLFSKSGIMSMCKAHLEDKYRFLKQVASTGFCQDORRLLGLLHEAIOIPROLGEVA SFSGSNIETSVSRSCFOFANNKPEIAALFLDWMRLPQSLVWMPVLRHRAAETAKHQ AKCNICEKPIIGFIRYSLKHFNYDVCQSCFFSGRTAKGKMHYPMVEYCTPTTSGED VRDFAKVLKKNKFTKRYFAKHPRMGYLPVQTVLEQDNLETPTVLINFPVVDYEPASSP QLSHDDTHSRIEHLRLAEMENRNGSYLNDISIPNESIDDEHLLIHOYQCSLNQESP LSQPRSPAQILISLESEERLELADLEDENRNLOSEYKELKQOHDHGLSPLPSP PEMPPISPQSPRDAELIAEAKLROHKGRLAEMQILEDNKQLESQHLRLQLLEQF QAEVRVNGTSVSSPSTSSQSDSSQPVLLHGVGSQTSIGILGEDDLLSAPQSTSELED VMEQLSSSFSPSSQRRDTPGKMKVEPVM"

BASE COUNT 773 a 618 c 646 g 617 t

Query Match 35.1%; Score 715.6; DB 5; Length 2654;

Best Local Similarity 71.3%; Pred. No. 3.8e-170;

Matches 943; Conservative 0; Mismatches 379; Indels 0; Gaps 0;

QY 719 AGTTCTGACCACTGGAAGCGTCTGCACCTTTCTCTGCAGGAAGTCTTGGTGTGGCTACAG 778

Db 3 AGTGGTGAGCAGTGAAGCGGCTGCAGATCTCCCTCCAGGACTTCTGACGTGGATGAAT 62

QY 779 CTGAAAGATGATGAATTAAGCCGCGAGGACCTATTGGAGGCGACTTTCCAGCAGTTTCA 838

Db 63 TTGAAGAACGATGAACCTAAGGAGGCAATGCCCCATTGGTGGTGATGCTCCACAGTCTGT 122

QY 839 AAGCAGAACGATGTACATAGGGCCCTTCAAGAGGGAATTTGAAAGAACTAAAGAACCTGTAATC 898

Db 123 CAGCAGAACGATGTTCACCGAATTTTCAAGCGAGAGTTGAAGGCAAGGAACCAAGTGGTC 182

QY 899 ATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAAGGACTAGAG 958

Db 183 ATGAGCGCTTTGGACACAGTGCATTTGTTCTTGGCTGATCCAGCAATCAGAGCTCCTGAA 242

QY 959 AAACCTCTACAGGAGCCAGAGAGCTGCCTCTCTGAGGAGAGAGAGCCCAAGATGTCACTCG 1018

Db 243 AGTCTTTTAACTGGACCAAG 302

QY 1019 CTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGAGAGAGAGAGAGAGAGAGAGAG 1078

Db 303 CGCATTCGGAAGTATGCCGAGGAGGTGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 362

QY 1079 GCTGACTGGCAG 1138

Db 363 GTTGATTGGCAG 422

QY 1654 ACTGCAGAAGGCCCTTGCTTGGATCTCTTGAGCCTGTGACCTGCATGTGATGCCTTGA 1713
Db 1182 ACTGCAGAAGGCCCTTGCTTGGATCTCTTGAGCCTGTGACCTGCATGTGATGCCTTGA 1241
QY 1714 CCAGCACAACTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTT 1773
Db 1242 CCAGCACAACTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTT 1301
QY 1774 GACCACATATTATGACCGCCTGGAGCAAGAGCACAACAAATTTGGTCAACGTCCTCTCTG 1833
Db 1302 GACCACATATTATGACCGCCTGGAGCAAGAGCACAACAAATTTGGTCAACGTCCTCTCTG 1361
QY 1834 CGTGGATATGTCTGAACTGGCTGCTGAATGTTTATGATACGGGACGACAGGGAGGAT 1893
Db 1362 CGTGGATATGTCTGAACTGGCTGCTGAATGTTTATGATACGGGACGACAGGGAGGAT 1421
QY 1894 CCGTGTCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAA 1953
Db 1422 CCGTGTCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAA 1481
QY 1954 GTACAGATACCTTTTCAAGCAAGTGGCAAGTTCACAGGATTTTGTGACCGCGCAGGCT 2013
Db 1482 GTACAGATACCTTTTCAAGCAAGTGGCAAGTTCACAGGATTTTGTGACCGCGCAGGCT 1541
QY 2014 GGGCCTCCTTCTGCATGATTCTATCCAA 2041
Db 1542 GGGCCTCCTTCTGCATGATTCTATCCAA 1569

AF339031 3521 bp mRNA linear VRT 20-APR-2001
Danio rerio dystrophin (dmd) mRNA, partial cds, alternatively spliced.
AF339031
AF339031.1 GI:13699249

Danio rerio.
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi;
Cypriniformes; Cyprinidae; Danio.
1 (bases 1 to 3521)
Bolanos-Jimenez,F., Bordais,A., Behra,M., Strahle,U., Sahel,J. and Rendon,A.
Dystrophin and Dp71, two products of the DMD gene, show a different pattern of expression during embryonic development in zebrafish Mech. Dev. 102 (1-2), 239-241 (2001)
21184125
11287201
2 (bases 1 to 3521)
Bolanos-Jimenez,F., Rendon,A. and Strahle,U.
Direct Submission
Submitted (22-JAN-2001) Laboratoire de Physiopathologie Retinienne, EMI 99-18 INSERM-Universite Louis Pasteur, 1, Place de l'Hopital, Strasbourg 67091, France
Location/Qualifiers
1. .3521
/organism="Danio rerio"
/db_xref="taxon:7955"
/chromosome="1"
/map="between z5508 and z5058"
<1. .3521
/gene="dmd"
<1. .3292
/gene="dmd"
/note="alternatively spliced"
/codon_start=2
/product="dystrophin"
/protein_id="AAK38376.1"
/db_xref="GI:13699250"
/translation="RPCRHWEDSHAKLTARVLTQLNMYKDSWDLEARKRVEPLIKKA

NEKLESWKVKVSHVEDLKGNADVQKLSKDLQWQTMNVNLANLKLTLYADDDTS
KVQMTESMNLAWANIKKRAGDEADLEAGRLQHQYVLDLEKFLNWLTEATTANVL
QDATFKEGLLLENPATVRLHLEQWQDLQAEIDAHRETYHSLDENGHRIVSSLEGTDNV
VLQKRLDDMGQRWHELNCNKVMSIRPYLDAGVDWKLHMSLQELLNLWLQKREELEKQ
KPVGGDVPTVHQOLLTHKAFRRELGAKEPVINGTLDNAKTFLAEMPREGLKQRPQKQD
VSPERVQNVGRILRKEVEDVTVRWKNLGAASVDWQQLLELRLMELODAQDQLDY
KLQAESEVKNWKPVGELLVDDLQNHIDRVKAFQEEIAPIDNVNHNVNQLASTFRPSD
IQLSPDNLSRIDDLNMRWRLQISIEEHLSQLTAFKDLGPSQNFHLASVESPLERSI
SPNNVPYYINHQTQTCWDHPKMAELYQSLADLNNVRFSAIRYRTAMKLRMMQKALCLDL
LSMPAAACEAFEQHNLKQNEQFMDIVQVINCILTSYDRLEQQHSSLVNPLCVDMLCNW
LLNVYDTGRAGKIRTLSEFKTGIISLCAHLEDKRYRFLFREVASATGFCDDQRRLLGLLH
DAIQIPROLGEVASFGGSNIEPSVRSQFQFANNKPELEASVFLDWMRLEPQSMVWLPV
LHRVAAEATAKHQAKCNICEKPIIGFRYRSKHFNYDICQSCFFSGRVAKGHKMOYP
MVEYCTPTTSGEDVRDFAKVLKNKFRKRYFAKHPRMGYLPVQVILEGDNMETPVTLI
NFWPVDHPPASSPOLSHDDTHSRTEHYASRDEHLLLIQHYCQSLNQGSPLSPQSPQAQ
ILISMETEEKGELEVLNDLEQENRKLQAEYDRLLKAHDHKGSLSPSPQMLPVSPQ
SPRDAELIAEAKLLRQHKRLEARMQILEHDKHQLKESQLTRLRLQLEQTESKVGNTAL
SSPSTASPRSDTSLASLRVAASQTETMGDDDELSSPTQDASTGLEDVIEQLNNSFPHS
QGGGRLLNP"
3293. .3521
3'UTR
BASE COUNT 1005 a 892 g 756 t
ORIGIN /gene="dmd"

Query Match 29.3%; Score 598.2; DB 5; Length 3521;
Best Local Similarity 64.7%; Pred. No. 2.1e-140;
Matches 923; Conservative 0; Mismatches 498; Indels 6; Gaps 2;

QY 614 GGAGATCAGCGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGGAGATCGATGGCA 673
Db 638 GGAACGGACAATGCTGTGTGCTACAAAAACGGCTTGATGACATGGGGCAGCGCTGGCAT 697
QY 674 AACATCTGTAGATGGACAGAACCGCTGGGTCTTTTACAAGACAGTTCTGACCAGTGG 733
Db 698 GAGTTGTGCAACAAAGTTATGAGTATAAGGCCCTATCTAGATGCCGGCGTGTGATGAGT 757
QY 734 AAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGTGGCTACAGCTGAAGATGATGAA 793
Db 758 AAACACTTGCACATGCTCTTACAAGAGCTGCTCAACTGGCTGACGCTGAAGAGGGAAGAG 817
QY 794 TTAAGCCGGCAGGCACCTATTGAGGGCGACTTTCCAGCAGCTTCAGAACGACAGACGATGTA 853
Db 818 CTGGAGAAAGCAGAAAGCCACTAGGGGGCGACGTGCCGACCGCTTACCAACAACCTCCTCACG 877
QY 854 CATAGGGCCCTTCAAGAGGGAATTGAAACCTAAGAACCTGTATCATGACTACTCTTGAG 913
Db 878 CACAAGGCTTTCAGAAAGAGAGCTGGGTGCCAAAGAACCTGTATCAATGGAACCTCTCGAC 937
QY 914 ACTGTACGAATATTCTTGACAGAGCAGCCCTTTGGAAGGACTAGAGAAACCTTACCAGGAG 973
Db 938 AATCGGAAAACCTTCTTGTGAGATGCCTCGTGAGGCGCT---GAAACAGAGACCTTGA 994
QY 974 CCCAGAGAGCTGCCTCCTGAGGAGAGAGAGCCAGAAATGTCACCTCGGCTTCTACGAAAGCAG 1033
Db 995 CAAAAGATGTAAGTCTCTGAGGAGCGGGTTCAAACACGTGGCGCCGCATATTGCGTAAAGAA 1054
QY 1034 GCTGAGGAGGTCAATACTGAGTGGGAAAAAATTGAACCTGCACCTCCGCTGACTGGCAGAGA 1093
Db 1055 GTGGAGGATGTAACGGTGCAGATGGAAGAATCTGGGTGCAGCTTCTGTGGACTGGCAGCAA 1114
QY 1094 AAAATAGATGAGACCTTGAAGAGACTCCAGGAACCTTCAAGAGGCCACGGATGAGCTGGAC 1153
Db 1115 CAGCTGGAACCTGGCCTTGGAGAGGCTGATGGAGCTCCAGGACGCCAGGATCAGCTGGAC 1174
QY 1154 CTCAAGCTGCCCAAGCTGAGGTGATCATCAAGGATCCTTGGCAGCCCGTGGCGGATCTCCTC 1213
Db 1175 TACAAGCTACGACAGGCTGAGAGCGGTGAAGAATCTTGAAGCCTGTTGGGAGCTGCTT 1234
QY 1214 ATTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAATTGCGCCT 1273
Db 1235 GTAGACGATTTGCAAAACCACATAGACAGAGTGAAGGCATTTTCAGGAAGAGATAGCTCCC 1294
QY 1274 CTGAAAGAGAAACGTGAGCCACCGTCAATGACCTTGCTGCCAGCTTACCACCTTTGGGCATT 1333

QY 601 TGATGAATCTAGTGGAGATCAGCAACTGCTGCTTTGGAAGAACAACCTTAAGGTATTGGG 660
Db 662 TGATGAATCCAGCGGTGATCATGCAACAGCTGCTTTGGAAGAACAACCTTAAGGTACTGGG 721
QY 661 AGATCGATGGGCAACATCTGTAGATGGACAGAACCGCTGGGTTCTTTTACAAGACAG 720
Db 722 AGATCGATGGGCAACATCTGCAGATGGACTGAAGACCGCTGGATTGTTTACAAGATAT 781
QY 721 TTCTGACCAGTGGGAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGTGGCTACAGCT 780
Db 782 TCTTCTAAATGGCAGCATTTTACTGAAGAACAGTGCCTTTTGTAGTACATGGCTTTCAGA 841
QY 781 GAAAGATGATGAATTA 797
Db 842 AAAAGAAGATGCAATGA 858

RESULT 9
MMGUTRPH
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
REMARK

REFERENCE
AUTHORS
TITLE
JOURNAL

FEATURES
source

1. .3161
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="AU#3"
/tissue_type="brain"
/clone_lib="adult mouse brain"
/dev_stage="adult"
1. .197
198. .3161
/function="Unknown"
/standard_name="G-utrophin"
/citation=[1]
/codon_start=1
/evidence=experimental
/product="G-utrophin (predicted protein)"
/protein_id="CAA58496.1"
/db_xref="GI:793852"
/db_xref="MGD:104631"
/db_xref="SPTREMBL:Q61636"
/translation="MQILRLQKCGKLMMAVVRTSLQKVVVLLHRLQRMVSSPRVQ
KICKIDIAEIDAHDNDFKSIDGNRQKMKVKGALGNSEEAATMLQHRLLDMNQNRWDLKAKS
ASIRAHLEASAEKNRLLASLEELIKWLNKMDLEELKKQMPIGGDVPAALQYDHCKVL
RRELKEYSVLNAVDAQRVFLADQPIEAPPEPRRNQPSKTELTPEERAQKIAKAMRK
QSSEVREKWNENLNAVTSNQKQVGRKALEKLRDLQGMDDLDADMKEAVEVRNGWKPVG
DLLIDSLODHIKTLAFREEIAPINLKVTMNDLSLSPLDLHPSLKMSRQLDLNM
RWKLLQVSVDDRLLKQLOEAHRDFGSSQHFSTSVQLPWQRSISHNKVPYYINHOTQT
TCWDHPKMTFLFQSLADLNNVRFSAYRTAIKRRLOKALCLDLLELNTTNEVFKQHL
NQNDQLLSVPDVINCLTFTYDGLQHLKDLVNVPLCVDMCLNLLNVLNVDYDGTGKIRV
QSLKIGLSLSKGLLEEKYRCLFKEVAGPTMCDQRQLGLLLHDAIQIPRLGEVAAF

GGSNIEPSVRSFCQQNNKPEISVKEFIDWMHLEPQSMVLPVLHRVAAAETAKHOAK
CNICECPVIGFRYSLKHFNDVQCSCFFSGRTAKGHLHYPMVEYCIPTTSGEDVR
DFTKVLKNKFRSKYFAKHPRGLYLPVQTVLEGDNLLETPTILISMWPEHYDPSQSPQL
FHDDTHSRIEQYATRLAOMERTNGSFLTDSSTTGSVEDEHALIQYCCOTLGGESPVS
QPQSPAQILKSVEREERELERIIADLEEQRNLQVEYEQHLRRLGLPVGSPDPS
IVSPHSTSESELIAEAKLLRQHKGRLEARMQILEDNKQLESQHLRLQLEQPDSD
SRINGVSPWASPOHSALSYSLDTPGQPFHQAAASEDLLAPPHDTSTDLTDVMEQINST
FPSCSSNVPSRPOAM"

BASE COUNT 887 a 784 c 802 g 688 t
ORIGIN

Query Match 27.1%; Score 554; DB 10; Length 3161;
Best Local Similarity 63.8%; Pred. No. 3.2e-129;
Matches 861; Conservative 0; Mismatches 480; Indels 9; Gaps 1;

QY 700 CTGGGTTCTTTTACAAGACAGTCTTGACCAGTGAAGCGTCTGCACCTTTCTCTGCAGGA 759
Db 512 CAGGGCCCATTTGGAGGCCAGTCTGAGAAATGAACCGGTTGCTGGCATCGCTGGAAGA 571
QY 760 ACTTCTGGTGTGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGACCTATTGGAGG 819
Db 572 GCTGATCAAAATGGCTCAATATGAAGATGAGGAGCTTAAGAAGCAGATGCCCATTTGGAGG 631
QY 820 CGACTTTTCCAGCAGTTCAGAAAGCAGAACGATGTACATAGGGCCTTCAAGAGGGAATTGAA 879
Db 632 GGACGTCCCTGCTTACAGCTCCAGTATGACCAGTGAAGGTGCTGAGACGTGAGCTAAA 691
QY 880 AACTAAAGAACCTGTAAATCATGAGTACTCTTGAGACTGTAGAAATATTTCTGACAGAGCA 939
Db 692 GGAGAAAGAGTATTCTGTGTGAACCGGTAGATCAAGCTCGAGTTTCTGGCTGATCA 751
QY 940 GCCTTT-----GGAAGGACTAGAGAACTCTACAGGAGCCCGCAGAGAGCTGCCCTCC 990
Db 752 GCCAATAGAGGCCCCCGGAAGAACCAAGAAAGCAATCAAGACAGAGTTGACTCC 811
QY 991 TGAGGAGAGAGCCCAAGAATGTCACTCGGCTTCTAGAAAGCAGGCTGAGGAGGTCAATAC 1050
Db 812 TGAGGAGAGAGCCCAAGAAGATCGCCAAAGCCATGCGCAAGCAGTCTTCTGAAGTCCGAGA 871
QY 1051 TGAGTGGGAAAATTGAACCTGCACCTCGCTGACTGCGCAGAGAAAATAGATGAGACCCCT 1110
Db 872 GAAAGTGGAAAATCTAAATGCTGCTACTAGCAACTAGCAAAAGCAAGTAGGGAAGGCGGT 931
QY 1111 TGAAAGACTCCAGGAACCTTCAAGAGGCGCACGAGTGAAGTGGACCTCAAGCTCGGCCAAGC 1170
Db 932 AGAGAAACTCCGAGAACCTGCAGGAGCTATGGACGACCTGGACGCAGACATGAAGGAGGT 991
QY 1171 TGAGGTGATCAAGGGATCCTGGCAGCCCGCTGGCGGATCTCTCTCATTTGACTCTCTCCAAGA 1230
Db 992 GGAGGCTGTGCGGAATGGCTGGAAGCCCGTGGGAGACCTGCTTATAGACTCCCTGCGAGGA 1051
QY 1231 TCACCTCGAGAAAAGTCAAGGCACCTTCGAGGAGAAAATTCGCGCCTCTGAAAGAGAACGTGAG 1290
Db 1052 TCACATCGAGAAAACCCCTGGCGTTTAGAGAGAAAATTCACCAATCAACTTAAAGATAA 1111
QY 1291 CCACGTCAATGACCTTGCTCGCCAGCTTACCACCTTTGGGCATTCAGCTCTCACCGTATAA 1350
Db 1112 AACAATGAATGACCTGTCCAGTCAGCTGTCTCCACTTGACTTGCATCCATCTCTAAAGAT 1171
QY 1351 CCTCAGCACTCTGGAAGACCTGAACACCCAGATGGAAGCTTTCGAGGTGGCCGCTCGAGGA 1410
Db 1172 GTCTCGCCAGCTGGATGACCTTAATATGCGATGGAAACTTCTACAGGTTTCCGTGGACGA 1231
QY 1411 CCGAGTCAGCAGCTGCATGAAGCCACACAGGACCTTGGTCCAGCATCTCAGCACTTTCT 1470
Db 1232 TCGCCTTAAGCAGCTCCAGGAAGCCACAGAGATTTTGGGCCATCTTCTCAACACTTTCT 1291
QY 1471 TTCCAGCTCTGTCCAGGTCCTCGGAGAGAGCCATCTCGCCAAAACAAAGTGCCTACTA 1530
Db 1292 GTCCACTTCAGTCCAGCTGCCGTGGCAGAGATCCATTTACATAATAAAGTGCCTATTA 1351
QY 1531 TATCAACACGAGACTCAAAACAACCTTGTGGAGCATCCCAAAATGACAGAGCTCTACCA 1590

/note="DRP2; membrane-associated cytoskeletal protein; contains two spectrin repeats; WW domain; Z2 domain"
/codon_start=1
/product="dystrophin-related protein 2 B-form splice variant"
/protein_id="AAG28485.1"
/db_xref="GI:11066167"
/translation="MQPLVMQGCPTLPRCHEWHAADRFHSSSLRNTCPQOVRAAV
TIPAPPWDGADPCLSPKLLGSGVAVGPLEPSANMLCWEIKKSHNLRLARLEAFSD
HSGKLQPLQEIIDWLSQKDEELSAQLPQGDVALVQVEKETHAAFEVSKGPYIY
SVLESAQFLSQHPFEELESHSEKDTSPRQIONLSRFVWKQATVASELWEKLTAR
CVDQHRHIEHTLEHLEIQGMEELSSLTQAEQVRAWEPIDGLFIDSLPEHQAIAK
LFKEEFSPVKDGVKLVNDLAHQLAISDVHLMSNSRALFQINVRWKQLQVSAERLQK
LQDAHRDFGPGSQHFLSTSVQVPWERAIISPKNVPIYINHQATTCWDHPKMTELYQTL
ADLNKIFSAYRTAMKLRVQKALRLDLVTLTTFNEDHDLQASEHVMVDVVEIHC
LTALYERLEEEERILVNPCLVDMSLNLNFDGSRGSKMRALSFKTGIACLCGTEV
KEKLYLFSQVANSKCDQRHLGALLHEAIOVPRQLGEVAFAFGGSNVEPSVRSQFRF
STGKPVIEASQFLEWVNLPEQSMVWLAVLHRVTVAEQVKHQTKCSICROCPKIFRYR
SLKQFNVDICQTCFLTGRASGNKLHYPIMEYTYPTTSSENMRDFAATLKNKFRSKQY
FSKHPQRGYLPVQSVLESQDSETPASSPMLPHADTHSRHFAASRLAEMESQNCFFN
DSLSRDDSDIEDQYLLRHSSPITDREPAGQQOAPCSMATSEKSGLEKILAHLEDENRI
LQGLRLRLKWOHEEAVEAPTLAGSAEATPDHRNEELAEARILRHQKSRLETRMQIL
EDHNKQLESQRLRLRELLQPTESDNGSAGSSLASPROSESGSHPREKGTTPDTE
EDHNKQLESQRLRLRELLQPTESDNGSAGSSLASPROSESGSHPREKGTTPDTE
AADDVGSQSDVSHSLEDIMEKLRHAFPPSVRSDDVTANTLLAS"
BASE COUNT 779 a 843 c 790 g 659 t 2 others

Query Match 25.9%; Score 528.4; DB 10; Length 3073;
Best Local Similarity 62.0%; Pred. No. 9.8e-123;
Matches 835; Conservative 0; Mismatches 511; Indels 0; Gaps 0;

QY 696 ACCGCTGGGTTCTTTTACAAGACAGTTCTGACACAGTGGAAAGCGTCTGCACCTTCTCTGC 755
DB 474 ACCTCCGCGCTCGCCTAGAGGCCTTCTCAGACCACACAGTGGAAAACTTCAGCTCCCACTTC 533
QY 756 AGGAACCTTCTGGTGTGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTG 815
DB 534 AAGAGATTATTGACTGGCTCAGCCAAAGGATGAAGAAATTCTCAGCTCAGCTGCCCTTGC 593
QY 816 GAGGCGACTTTCAGGAGTTCAGAAAGCAGACGATGTACATAGGCGCTTCAAGAGGGAAT 875
DB 594 AAGGGGATGTGGCCCTGGTACACAGGAGGAGACACATGCTGCTCTTATGGAAGAAG 653
QY 876 TGAAGAACTAAAGAACCTGTATCATGAGTACTCTTGAGACTGTACGAATATTCTTGACAG 935
DB 654 TGAAGTCTAAAGGCCCTATATCTACTCTGTGCTGGAATCGGCACACAGGCTTCTCTGTCTC 713
QY 936 AGCAGCTTTTGAAGGACTAGAGAAACTCTACCAGGAGCCAGAGAGCTGCTCTCTGAGG 995
DB 714 AGCAGCCATTTGAGGAATTAGAGGAGTCTCATTTCTGAGAGCAAGATACCTCCCAAGAC 773
QY 996 AGAGAGCCCAAGATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGT 1055
DB 774 AGCGGATTCAAAACCTCAGTCGCTTTGTGTGGAAGCAGGCAACAGTGGCCCACTGAACTGT 833
QY 1056 GGGAAAAATTGAACCTGCACTCCGCTGACTGGCCAGAGAAAAATAGATGAGACCTTGAAA 1115
DB 834 GGGAGAGAGCTGACAGCCCGCTGTGTGGATCAGCATCGCCACATTGAGCACACCTTGAAC 893
QY 1116 GACTCCAGGAAGTCAAGAGGCGCCAGGATGAGTGGACCTCAAGCTGCGCCCAAGCTGAGG 1175
DB 894 ATCTATTGGAGATCCAAAGGGCAATTGAGGAACTGAGCAGTACTTTGACCCCAAGCAGAGG 953
QY 1176 TGATCAAGGGATCTCTGGCAGCCCGTGGCGGATCTCTCTCACTTACTCTCTCAAGATCACC 1235
DB 954 GAGTCCGAGCCACATGGGAGCCCATAGGAGATCTCTTATCGATTCCCTCCAGAGGCATA 1013
QY 1236 TCGAGAAAGTCAAGGCACTTCGAGGAGAAATTCGCCCTCTGAAAGAGAACGTGAGCCACG 1295
DB 1014 TCCAAGCCATCAAGTTATTCAAGAAGAAGAAATCTCTCTCTGTGAAAGACGGGTGAAGTAG 1073
QY 1296 TCAATGACCTTGCTCGCCAGCTTACCACTTTTGGGCAATTCAGCTCTCACCGTATAACCTCA 1355

Db 1352 CATCAACCATCAAAACACAGACAAACCTGTTGGGATCATCTTAAATGACTGAGCTCTTCCA 1411
QY 1591 GTCTTTAGCTGACCTGAATAATGTGACGATTTCTCAGCTTATAGGACTGCCATGAAACTCCG 1650
Db 1412 ATCCCTTGCTGATCTGAATAATGTACGTTTCTCTGCCTACCGACAGCAATCAAAATTCG 1471
QY 1651 AAGACTGCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCCTGTGACGCTGCATGTGATGCCCTT 1710
Db 1472 AAGGCTGCAAAAAGCATTATGCTGGATCTCTTAGAGCTGAATACGACGAATGAAGTTT 1531
QY 1711 GGACGAGCACAAACCTCAAGCAAAATGACGACGCCCATGGATATCCTGCGAGATTATTAATTG 1770
Db 1532 CAAGCAGCACAAACCTGAACCAAAATGATCAGCTCCTGAGTGTCCGACGCTCATCAACTG 1591
QY 1771 TTTGACCACTATTATGACCGCTGGAGCAAGAGCACAACAATTTGGTCAACGTCCTCT 1830
Db 1592 TCTGACCACTTACGATGAGTGGCTTGAGCAGCTGCACAAGGACTTGGTCAATGTTCCACT 1651
QY 1831 CTGCGTGGATATGTCTGAACCTGGCTGGCTGAATGTTTATGATACGGGACGACAGGGAG 1890
Db 1652 CTGCGTGGATATGTCTCAACTGGCTGGCTCAACGATATACGACACGGGCGGAGTGGAAA 1711
QY 1891 GATCCGCTGCTCTCTTTTAAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGA 1950
Db 1712 AATTCGGGTACAGAGTCTGAAGATTGGATTGATGATGCTCTCTCCAAAGGCTCTTAGAAGA 1771
QY 1951 CAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCACAGGATTTTGTGACCGCGCAG 2010
Db 1772 GAAATACAGATGTCTCTTAAAGGAGGTGGCAGGGGCAACTGAGATGTGTGACCGCGCA 1831
QY 2011 GCTGGGCTCTCTTCTGCATGATTTCTATCCA 2040
Db 1832 GCTGGGCTGCTACTTACGATGCCATCCA 1861

RESULT 10
AF195788
LOCUS
DEFINITION
Rattus norvegicus dystrophin-related protein 2 B-form splice variant (Drp2) mRNA, complete cds.
ACCESSION
AF195788
VERSION
AF195788.1
KEYWORDS
Rattus norvegicus.
SOURCE
Rattus norvegicus.
ORGANISM
Rattus norvegicus.
REFERENCE
1 (bases 1 to 3073)
AUTHORS
Roberts,R.G. and Sheng,M.
TITLE
Association of dystrophin-related protein 2 (DRP2) with postsynaptic densities in rat brain
JOURNAL
Mol. Cell. Neurosci. 16 (5), 674-685 (2000)
MEDLINE
20538711
PUBMED
11083927
REFERENCE
2 (bases 1 to 3073)
AUTHORS
Roberts,R.G. and Sheng,M.
TITLE
Direct Submission
JOURNAL
Submitted (18-OCT-1999) Division of Medical and Molecular Genetics, GKT Medical School, 8th Floor, Guy's Hospital, Guy's Tower, London SE1 9RT, UK
FEATURES
Location/Qualifiers
source
1. .3073
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
1. .3073
/gene="Drp2"
69. .70
/gene="Drp2"
/note="splice acceptor AG dinucleotide used by A-form"
200. .3073
/gene="Drp2"
CDS

Db 1074 TGAATGACCTGGCCACCAGCTTGCTATTCTGATGTGCACTTGTCATGGAGAATTCAA 1133
QY GCATCTGGAAGACCTGAACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAG 1415
Db 1134 GGGCTCTGGAACAGATCAACGCTCGCTGGAAACAGCTCCAGGTGTGCTGAGAGGC 1193
QY 1416 TCAGGAGCTGCATGAAGCCACAGGGACTTTGGTCCAGCATCTCAGCACCTTCTTTCCA 1475
Db 1194 TTAACAGCTCCAGGATGCCCCAGGACTTTGGCCCTGGGTACAGCACCTTCTCTCCA 1253
QY 1476 CGTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAGTGGCCCTACTATATCA 1535
Db 1254 CTCTGTTCAGGTTCCCTGGGAAAGAGCAATTCTCCCAATAAAGTCCCTACTACATCA 1313
QY 1536 ACCACGAGACTCAACAACACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACAGTCTT 1595
Db 1314 ACCACAGGCTCAGACCACATGCTGGGACCATCCCAAGATGACCGAGTTATACCAAAACC 1373
QY 1596 TAGCTGACCTGAATAATGTGAGATTTCTCAGCTTATAGGACTGCCATGAACTCCGAAGAC 1655
Db 1374 TAGCTGACCTGAACAACATCAAGTTCTCAGCTTACCGCACTGCCATGAAGTCCGCCGAG 1433
QY 1656 TGCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTGAGCTGCATGTGATGCCCTCTCTGCG 1715
Db 1434 TCCAGAAGGCCCTGCGCTTGGATCTGGTAACTTTAACTACAGCTCTGGAGATCTTCAATG 1493
QY 1716 AGCACAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTGA 1775
Db 1494 AGCAGGACTTGCAGGCCAGTGAACATGTAATGGATGTGGTGGAGTCAATCACTGCTTGA 1553
QY 1776 CCACTATTTATGACCGCCTGGAGCAAGAGCACAACAATTTGGTCAACGTCCCTCTCTGCG 1835
Db 1554 CTGCCTTGATGAACGACTGGAGGAGGAAAGAGGCATCCTGTCAACGTGCCGTGTGTG 1613
QY 1836 TGGATATGTGTGTAAGTGGCTGTGTAATGTTATGACGGGACGAACAGGGAGGATCC 1895
Db 1614 TAGACATGAGCCTCAACTGGCTCTCAACGTTTTTIGATAGTGGCCGCAAGTGGAAGATGC 1673
QY 1896 GTGTCCTGTCTTTAAACTGGCATCATTTCCCTGTGTAAGCAGACATTTGGAAGACAAGT 1955
Db 1674 GAGCATGTCTTTAAGACTGGCATCGCATGCTGTGGCACTGAAGTGAAGAAAAAC 1733
QY 1956 ACAGATACCTTTTCAAGCAAGTGGCAAGTTTCAACAGGATTTTGTGACCAGCGCAGGCTGG 2015
Db 1734 TTCAGTATCTCTTCAAGCAAGTGGCCCAATTCAGGCAGCAAGTGTGACCAGCGCCACCTCG 1793
2016 GCCTCCTTCTGCATGATTTCTATCCAA 2041
Db 1794 GTGCCCTGCTTCATGAGCCATCCAA 1819

RESULT 11
AF195787
LOCUS 3140 bp mRNA linear ROD 02-MAY-2001
DEFINITION Rattus norvegicus dystrophin-related protein 2 A-form splice variant (Drp2) mRNA, complete cds.
ACCESSION AF195787
VERSION AF195787.1 GI:11066164
KEYWORDS Rattus norvegicus.
SOURCE Rattus norvegicus
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE 1 (bases 1 to 3140)
AUTHORS Roberts,R.G. and Sheng,M.
TITLE Association of dystrophin-related protein 2 (Drp2) with postsynaptic densities in rat brain
JOURNAL Mol. Cell. Neurosci. 16 (5), 674-685 (2000)
MEDLINE 20538711
PUBMED 11083927
REFERENCE 2 (bases 1 to 3140)
AUTHORS Roberts,R.G. and Sheng,M.

TITLE Direct Submission
JOURNAL Submitted (18-OCT-1999) Division of Medical and Molecular Genetics, GKT Medical School, 8th Floor, Guy's Hospital, Guy's Tower, London SE1 9RT, UK
FEATURES
source Location/Qualifiers
1. .3140
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
1. .3140
/gene="Drp2"
267. .3140
/gene="Drp2"
/note="DRP2; membrane-associated cytoskeletal protein; contains two spectrin repeats; WW domain; ZZ domain"
/codon_start=1
/product="dystrophin-related protein 2 A-form splice variant"
/protein_id="AAG28484.1"
/db_xref="GI:11066165"

/translation="MQPLVMQGPCPYTLPRCHEWHAADRFHSSSLRNTCPQPQVRAAV TIPAPPWDGAGDPCLSPKLLNGSVGAVPLEPSAMNLWNEIKKKSHNLARLEAFSD HSKLQLPQEIIDWLSQKDEELSAQLPQGDVALVOQEKETHAAFMEEVKSQGYIY SVLESAQAFLSQHPFEELESHSESKDTSRQIQNLRSFVWKQATVASSELWEKLTAR CVDQHRHIEHTLEHLLEIQGAMEELSTLTQAEQVTRATEPFGDLFIDSLPEHQAIAK LFKKEESPVKDGVKLVNDLAHQLAISDVHLSMENSRALEQINVRWKQLQVSVARLKQ LQDAHRDPGPGSQHFLSTSVQVPWERAISPKNVPYINHQAOCTCWDHPKMTELYQTL ADLNNIKFSAYRTAMKLRVQKALRDLTLTALFEFNEHDLQASEHVMDVVEVIHC LTALYERLEEEERGLVNVPLCVDMSLWLLNVFDSRSGKMRALSFKTGIACLGCTEV KEKLYLFISOVANSKQDQRLHGLLHEAIQVPRQLGEVAAFGGSNVPSVRSCTFR STGKPVIEASQFLEWNLPEQSMVWLAVLHRVTVAEQVKHQTKCSICRQCPKGFYR SLKQFNVDICQTCFLTGRASKGNKLHYPIMEYTPPTSSENMRDFATTLKNKFRSKQY FSKHPQRGYLPVQSVLESDCSETPASSPMLPHADTHSRIEHFASRLAEMESQNSCFN DLSLSPDDSIDEDQYLLRHSSPITDREPFGQAAPCSMATSKESKLEKILAHLEDENRI LQELRLKQWHEEAVEAPTAEQSAEATPDHNEELAEARILRQHKSRLETRMQIL EDHNKQLESQDLRLRELLQPPPTESDNGSAGSSSLASPRQSEGSHPREKQGTTPDTE AADDVGSKSQDVSHSLEDIMEKLRHAFPSVRSSDVTANTLLAS"

BASE COUNT 790 a 864 c 794 g 690 t 2 others
ORIGIN

Query Match 25.9%; Score 528.4; DB 10; Length 3140;
Best Local Similarity 62.0%; Pred. No. 9.8e-123;
Matches 835; Conservative 0; Mismatches 511; Indels 0; Gaps 0;

QY 696 ACCGCTGGGTCTTTTACAAGACAGTTCTGACCAGTGAAGCGTGTGCACCTTTCTCTGC 755
Db 541 ACCTCCGCGCTCGCCTAGAGGCTTCTCAGACCACAGTGGAAAACTTCAGCTCCCACTTC 600
QY 756 AGGAACTTCTGGTGTGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTG 815
Db 601 AAGAGATTATTGACTGGCTCAGCCAAAAGGATGAAGAATTGTACAGTCACTGCCCCCTTC 660
QY 816 GAGGCGACTTTCAGCAGCTTCAGAAAGCAGAACGATGTACATAGGCGCTTCAAGAGGGAAT 875
Db 661 AAGGGGATGTGGCCCTGGTACAAACAGGAGAGGAGACACATGCTGCCTTTATGGAAGAAG 720
QY 876 TGAAAACTAAAGAACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAG 935
Db 721 TGAAGTCTAAAGGCCCTATATCTACTCTGTCTGGAATCGGCACAGGCTTTCTCTGCTC 780
QY 936 AGCAGCCTTTGGAAGGACTAGAGAAACTCTACCAGGAGCCAGAGAGCTGCCTCCTGAGG 995
Db 781 AGCACCCATTGAGGAATTAGAGGAGTCTCATTTCTGAGAGCAAGATACCTCCCCAAGAC 840
QY 996 AGAGAGCCCCAGAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGT 1055
Db 841 AGCGGATTCAAAACCTCAGTCGCTTGTGTGGAAGCAGGCAACAGTGGCCAGTGAACGTGT 900
QY 1056 GGGAAAAATTGAACCTGCACCTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGA 1115
Db 901 GGGAGAAGCTGACAGCCCGCTGTGTGGATCAGCATCGCCACATTTGAGCACACCCCTGGA 960
QY 1116 GACTCCAGGAACCTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGG 1175

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 263 to 3189)		Robert, R.G., Freeman, T.C., Kendall, E., Vetrie, D.L., Dixon, A.K., Shaw-Smith, C., Bone, Q. and Bobrow, M. Characterization of DRP2, a novel human dystrophin homologue Nat. Genet. 13 (2), 223-226 (1996)	
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
gene			
5'UTR			
CDS			
3'UTR			
BASE COUNT	856 a	964 c	913 g 766 t
ORIGIN			
Query Match	25.8%;	Score 527.4;	DB 9; Length 3499;
Best Local Similarity	62.0%;	Pred. No. 1.8e-122;	
Matches	834;	Conservative	0; Mismatches 511; Indels 0; Gaps 0;
Qy	696	ACCGCTGGGTTCTTTTACAAGACAGATTCTGACCAAGCGTGGAAAGCGTCTGCACCTTTCTCTGC	755
Db	593	ACCTCCCGCTCGCTAGAGGCCTTCTCAGACCCACAGTGGAAAGCTTCAGCTCCCTCTC	652
Qy	756	AGGAACCTTCTGGTACAGCTGAAAGATGATGAATTAAGCCGCGCAGGCACCTATTG	815
Db	653	AAGAGATTATTGACTGGCTCAGCCAAAAGGATGAGGAGTTGTCAGCTCAGCTGCCCCCTAC	712
Qy	816	GAGGCGACTTTCAGCAGTTCAGAAAGCAGAACGATGTACATAGGGCCCTTCAAGAGGGAAT	875
Db	713	AGGGGATGTGGCCCTGGTGCACACAGGAGAGACACATCGGGCCCTTATGGAAGAAG	772
Qy	876	TGAAAACTAAAGAACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAG	935
Db	773	TCAAGTCTCGGGGCCCTACATCTATTCTGTGTGGAGTCAAGTCAAGGCCTTCTGTCTCC	832


```
Db 193 GCAGAAAGCCCTTTGGATCTCTTGAGCCTGTGCAGCTGCATGTGATGCCCTTGACCA 252
2Y 1717 GCACAAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAAATTGTTGAC 1776
Db 253 GCACAAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAAATTGTTGAC 312
2Y 1777 CACTATTATGACCGCCTGGAGCAAGAGCAACAATTTGGTCAACGTCCTCTCTGCGT 1836
Db 313 CACTATTATGACCGCCTGGAGCAAGAGCAACAATTTGGTCAACGTCCTCTCTGCGT 372
2Y 1837 GGATATGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCG 1896
Db 373 GGATATGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCG 432
2Y 1897 TGTCCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTA 1956
Db 433 TGTCCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTA 492
2Y 1957 CAGATACCTTTTCAAGCAAGTGGCAAGTTTCAACAGGATTTGTGACCAGCGCAGGCTGGG 2016
Db 493 CAGATACCTTTTCAAGCAAGTGGCAAGTTTCAACAGGATTTGTGACCAGCGCAGGCTGGG 552
2Y 2017 CCTCCTTCTGCATGATTCTATCCAA 2041
Db 553 CCTCCTTCTGCATGATTCTATCCAA 577

RESULT 14
BC028720
LOCUS
DEFINITION
BC028720
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 4658)
Direct Submission
Submitted (29-APR-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki
Toshiyuki and Piero Carninci (RIKEN)
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
contact: amadan@systemsbiology.org
Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha
Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAC Plate: 46 Row: j Column: 3
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 5032296.
Location/Qualifiers
1. .4658
/organism="Homo sapiens"
/db_xref="LocusID:1756"
/db_xref="taxon:9606"
/clone="MGC:33083 IMAGE:4822807"
/tissue_type="Testis"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"

FEATURES
source
```

```

/note="Vector: pBluescript"
130. .2037
/codon_start=1
/product="dystrophin (muscular dystrophy, Duchenne and
Becker types)"
/protein_id="AAH28720.1"
/db_xref="GI:20379676"
/translation="MREQLKGHETQTCWDHPKMTELYQSLADLNNVRSAYRTAMKL
RRLQKALCLDLLSLSAACDALDQHNLKQNDQPMIDILQIINCLTTIYDRLEQHNNLVN
VPLCDVMCLNWLNVYDTRGTRIRVLSFKTGIIISLKAHLEDKYRLEKQVASSTGF
CDQRLGLLLHDSIQIPROLGEVASFSGSNIKCEPIIGFRYSLKHFNYDICQSCFFSG
LEPQSMVWLPVLRHVAATAETAKHQAKNICKEPIIGFRYSLKHFNYDICQSCFFSG
RVAKGHMHPMVEYCTPTTSGEDVRDEAKVKNKFKRTKRYFAKHPRMGVLPVQTVLE
GDNMETPVTLINFWPVDSPASSPQLSHDDTHSRIEHYASRLAEMENSNGSYLNDISI
PNESIDDEHLLIQHQCQLNQDPSLSPQSPRPAQILISLESEERGELELRLADLEEENR
NLQAEYDRLKQOHEHKGSLPLSPPEMPTSPQSPRDAELIAEAKLLRQHKRLEARM
QILEDHNKQLESQHLRLQLLEQPAEAKVNGTTVSSPSTLQSRSDSSOPMLLRVVG
QTSDSMGEEDLLSPPDQDSTSTGLEEVMEQLNNSFPSSRGNVGSLSFHMADDLGRAMESL
VSVMTDEEGAE"
BASE COUNT 1406 a 955 c 934 g 1363 t
ORIGIN
Query Match 24.7%; Score 505; DB 9; Length 4658;
Best Local Similarity 100.0%; Pred. No. 8.4e-117;
Matches 505; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1537 CCACGAGACTCAAAACAACCTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAAGTCTTT 1596
Db 150 CCACGAGACTCAAAACAACCTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAAGTCTTT 209
QY 1597 AGCTGACCTGAATAATGTGAGTCTCTGAGCTTATAGGACTGCCATGAAACTCCGAAGACT 1656
Db 210 AGCTGACCTGAATAATGTGAGTCTCTGAGCTTATAGGACTGCCATGAAACTCCGAAGACT 269
QY 1657 GCAGAAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTGAGCTGATGATGCTTGGACCA 1716
Db 270 GCAGAAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTGAGCTGATGATGCTTGGACCA 329
QY 1717 GCACAAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAAATTGTTGAC 1776
Db 330 GCACAAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAAATTGTTGAC 389
QY 1777 CACTATTTATGACCGCCTGGAGCAAGACACACAATTTGGTCAACGTCCTCTCTGCGT 1836
Db 390 CACTATTTATGACCGCCTGGAGCAAGACACACAATTTGGTCAACGTCCTCTCTGCGT 449
QY 1837 GGATATGTGTCTGAACCTGGCTGTGAATGTTTATGATACGGGACGACAGGGAGGATCCG 1896
Db 450 GGATATGTGTCTGAACCTGGCTGTGAATGTTTATGATACGGGACGACAGGGAGGATCCG 509
QY 1897 TGTCCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTA 1956
Db 510 TGTCCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTA 569
QY 1957 CAGATACCTTTTCAAGCAAGTGGCAAGTTTCAACAGGATTTTGTGACCAGCGCAGGCTGGG 2016
Db 570 CAGATACCTTTTCAAGCAAGTGGCAAGTTTCAACAGGATTTTGTGACCAGCGCAGGCTGGG 629
QY 2017 CCTCCTTCTGCATGATTCTATCCAA 2041
Db 630 CCTCCTTCTGCATGATTCTATCCAA 654

RESULT 15
HSDMDF1
LOCUS
DEFINITION
HSDMDF1 Human fetal mRNA fragment of DMD gene (DMD= Duchenne muscular
dystrophy).
ACCESSION X06179 M22819
VERSION X06179.1 GI:30839
KEYWORDS Duchenne muscular dystrophy.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
```


Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 630)
AUTHORS Cross,G.S., Speer,A., Rosenthal,A., Forrest,S.M., Smith,T.J.,
Edwards,Y., Flint,T., Hill,D. and Davies,K.E.
TITLE Deletions of fetal and adult muscle cDNA in Duchenne and Becker
muscular dystrophy patients
JOURNAL EMBO J. 6 (11), 3277-3283 (1987)
MEDLINE 88111512
PUBMED 3428261
COMMENT Data kindly reviewed (12.OCT.1988) by DAVIES K.E.
FEATURES
source

Location/Qualifiers

1..630

/organism="Homo sapiens"

/db_xref="taxon:9606"

/clone="Cf16."

/dev_stage="fetus"

1..630

/gene="DMD"

<1..>630

/gene="DMD"

/codon_start=2

/protein_id="CAA29545.1"

/db_xref="GI:30840"

/translation="SKLIGTKLSEDEETEVEQMNLLNSRWECLRVASMEKQSNLHR
VLMDLQNKQLKELNDWLTKEERTKMEEPGLPDLEDKRQVQOHKVLQEDLEQEQV
RVNSLTHMVVVDESQDHATAALEEQKLVGLDRWANICRWTEDEWVLLQDTLLKQWR
LTEEQCLFSAWLSEKEDAVNKIHTTGFKDQNEMLSSLQKLAVLKADLKK"

241 a 107 c 142 g 140 t

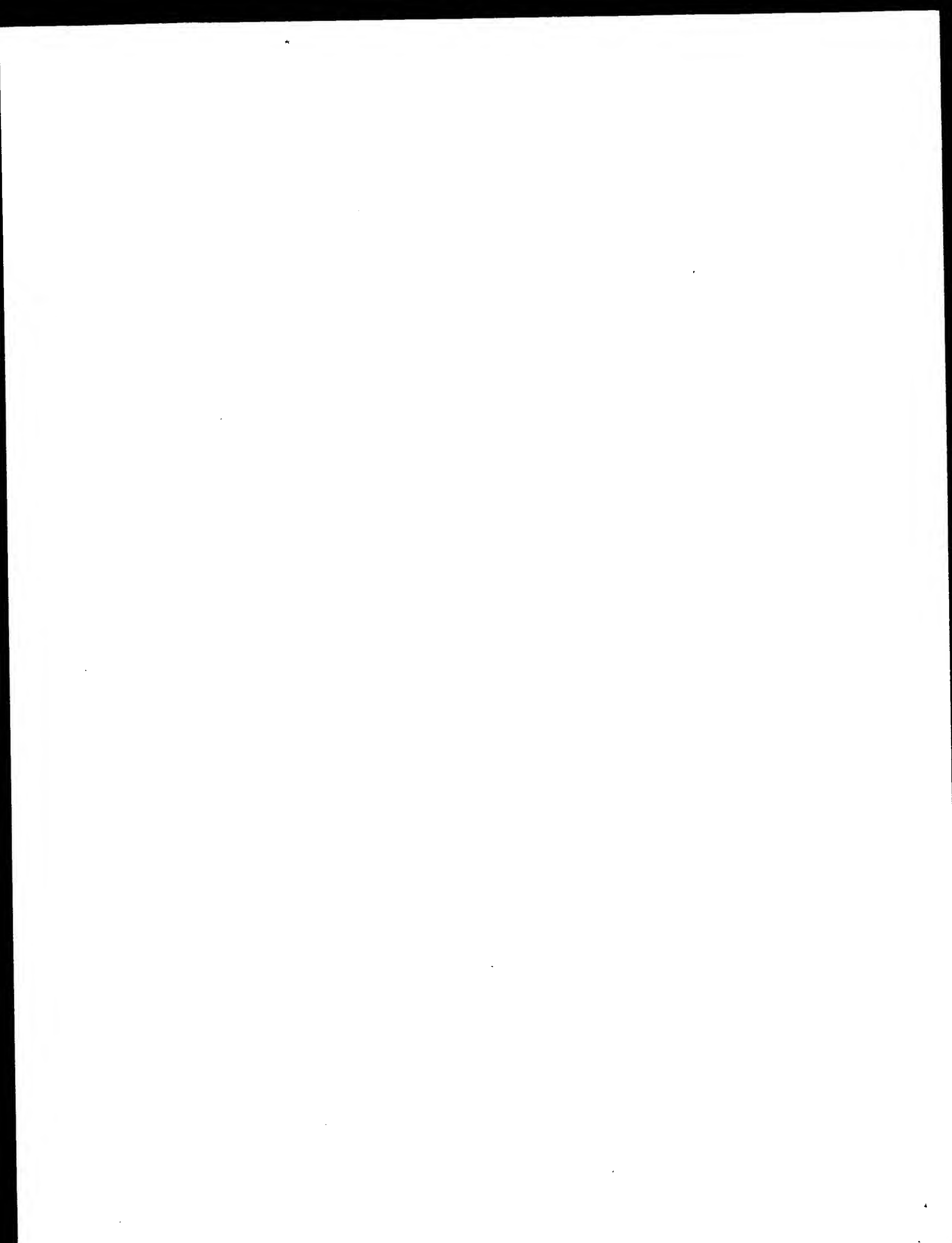
BASE COUNT

ORIGIN

Query Match 22.9%; Score 468; DB 9; Length 630;
Best Local Similarity 91.7%; Pred. No. 1.8e-107;
Matches 495; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 260 AGTAAGCTGATTGGAACAGGAAAATATCAGAAGATGAAGAACTGAAGTACAAGAGCAG 319
Db 2 AGTAAGCTGATTGGAACAGGAAAATATCAGAAGATGAAGAACTGAAGTACAAGAGCAG 61
QY 320 ATGAATCTCTAAATCAAGATGGGAATGCCTCAGGGTAGCTAGCATGGAACAAAGC 379
Db 62 ATGAATCTCTAAATCAAGATGGGAATGCCTCAGGGTAGCTAGCATGGAACAAAGC 121
QY 380 AATTACATAGAGTTTAATGGATCTCCAGAATCAGAACTGAAGAGTTGAATGACTGG 439
Db 122 AATTACATAGAGTTTAATGGATCTCCAGAATCAGAACTGAAGAGTTGAATGACTGG 181
QY 440 CTAACAAAACAGAAAGAAAGAAACAAACATAGGTCCTCAAGAGATCTAGAACAAAGAA 559
Db 182 CTAACAAAACAGAAAGAAAGAAAGAAAGAAATGGAGGAGAGCCCTCTTGGACCTGATCTT 241
QY 500 GAAGACCTAAAACGCCCAAGTACAACAAACATAGGTCCTCAAGAGATCTAGAACAAAGAA 559
Db 242 GAAGACCTAAAACGCCCAAGTACAACAAACATAGGTCCTCAAGAGATCTAGAACAAAGAA 301
QY 560 CAAGTCAGGTCATCTCTCACTCACATGGTGGTGGTAGTTGATGAATCTAGTGGAGAT 619
Db 302 CAAGTCAGGTCATCTCTCACTCACATGGTGGTGGTAGTTGATGAATCTAGTGGAGAT 361
QY 620 CACGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGGAGATCGATGGGCAACATC 679
Db 362 CACGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGGAGATCGATGGGCAACATC 421
QY 680 TGTAGATGGACAGAGACCGCTGGTCTTTTACAAGACAGATTCTGACCACTGGAAGCGT 739
Db 422 TGTAGATGGACAGAGACCGCTGGTCTTTTACAAGACAGATTCTGACCACTGGAAGCGT 481
QY 740 CTGCACCTTTCTCTGCAGGAACCTTCTGGTGTGGCTACAGCTGAAAGATGATGAATTAAGC 799
Db 482 CTTACTGAAGAACAGTGCCTTTTGTAGTGCATGGCTTTTCAGAAAAGAGATGCAGTGAAC 541

Search completed: April 25, 2003, 14:28:52
Job time : 3514.87 secs



GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 07:25:57 ; Search time 284.596 Seconds
(without alignments)
16150.348 Million cell updates/sec

Title: US-09-845-416-12_COPY_960_3000
Perfect score: 2041
Sequence: 1 tccttcacagcatttggaag.....ttctgcatgattctatccaa 2041

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4313688

Minimum DB seq length: 0
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*

1: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1980.DAT:*

2: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1981.DAT:*

3: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1982.DAT:*

4: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1983.DAT:*

5: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1984.DAT:*

6: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1985.DAT:*

7: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1986.DAT:*

8: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1987.DAT:*

9: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1988.DAT:*

10: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1989.DAT:*

11: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1990.DAT:*

12: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1991.DAT:*

13: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1992.DAT:*

14: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1993.DAT:*

15: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1994.DAT:*

16: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1995.DAT:*

17: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1996.DAT:*

18: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1997.DAT:*

19: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1998.DAT:*

20: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA1999.DAT:*

21: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA2000.DAT:*

22: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA2001A.DAT:*

23: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA2001B.DAT:*

24: /SIDS2/gcgdata/geneseq/geneseq-nbml/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2041	100.0	3510	24	Human dystrophin m
2	2041	100.0	4476	24	Adeno-associated v
3	1690.8	82.8	3531	24	Human dystrophin m
4	1690.8	82.8	4498	24	Adeno-associated v
5	1683	82.5	3858	24	Human dystrophin m
6	1683	82.5	4825	24	Adeno-associated v
7	1683	82.5	4848	24	Adeno-associated v
8	1581	77.5	4414	24	Adeno-associated v
9	1569	76.9	3446	24	Human dystrophin m

10	1542	75.6	3999	24	AAD37234	Human dystrophin m
11	1542	75.6	4966	24	AAD37256	Adeno-associated v
12	1542	75.6	4990	24	AAD37262	Adeno-associated v
13	1359	66.6	4182	24	AAD37230	Human dystrophin m
14	1328.8	65.1	2169	24	AAD37232	Human dystrophin r
15	1324	64.9	1821	24	AAD37241	Human dystrophin r
16	1257	61.6	4402	21	AAZ48567	A rod shortened dy
17	1136.6	55.7	4402	21	AAZ48568	A rod shortened dy
18	937	45.9	1434	24	AAD37243	Human dystrophin r
19	725.4	35.5	1991	24	AAD37231	Human dystrophin N
20	717	35.1	1667	24	AAD37235	Human dystrophin N
21	671.4	32.9	4075	21	AAZ48569	A rod shortened dy
22	667.2	32.7	3747	21	AAZ48566	A rod shortened dy
23	613.6	30.1	3163	21	AAZ48571	A rod shortened dy
24	574.6	28.2	3275	10	AAZ48571	Partial sequence o
25	394.8	19.3	3172	21	AAZ48570	A rod shortened dy
26	393	19.3	887	24	ABK81993	Human dystrophin C
27	390	19.1	1340	24	AAD37239	Human dystrophin N
28	387	19.0	387	24	ABK81990	Human dystrophin s
29	331	16.2	333	24	ABK81967	Human dystrophin s
30	327	16.0	327	24	ABK81966	Human dystrophin s
31	324	15.9	324	24	ABK81991	Human dystrophin s
32	215	10.5	215	24	ABK81992	Human dystrophin h
33	198.2	9.7	1500	22	AAF84672	Nucleotide sequenc
34	190.2	9.3	256	22	ABA69736	Human foetal liver
35	190.2	9.3	256	22	ABA36636	Probe #15102 for g
36	190.2	9.3	256	22	AAK17925	Human brain expres
37	190.2	9.3	256	22	AAK43799	Human bone marrow
38	190.2	9.3	256	22	AAI24574	Probe #14507 for g
39	190.2	9.3	256	22	AAI49820	Probe #18506 used
40	190.2	9.3	256	24	ABS18027	Human genome-deriv
41	177	8.7	466	22	ABA58628	Human foetal liver
42	177	8.7	466	22	ABA27626	Probe #6092 for ge
43	177	8.7	466	22	AAK06756	Human brain expres
44	177	8.7	466	22	AAK32467	Human bone marrow
45	177	8.7	466	22	AAI16084	Probe #6017 for ge

ALIGNMENTS

RESULT 1
AAD37240
ID AAD37240 standard; DNA; 3510 BP.
XX
AC AAD37240;
XX
DT 21-AUG-2002 (first entry)
XX
DE Human dystrophin minigene delta3510.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Homo sapiens.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
XX
PI XIAO X;
XX
DR WPI; 2002-049342/06.
XX
PT New dystrophin minigene for treating Duchenne or Becker muscular dystrophy comprises an N-terminal domain or modified N-terminal domain,

rod repeats, H1 and H4 domains and a cysteine rich domain of a dystrophin gene -

Example 1; Page 51-52; 71pp; English.

The present invention relates to an isolated nucleotide sequence encoding a dystrophin minigene. The minigene comprises N-terminal or modified N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control element. The dystrophin minigene in operable linkage with an expression control element, in a recombinant adeno-associated virus or retrovirus is useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is human dystrophin minigene delta3510 containing nucleotides 1-1668 (N-terminus, hinge H1 and rods R1 and R2), 8407-10227 (rods R23 and R24, hinge H4 and CR domain) and 11047-11058 (dystrophin last 3 amino acids).

Sequence 3510 BP; 1073 A; 787 C; 828 G; 822 T; 0 other;

Query Match 100.0%; Score 2041; DB 24; Length 3510; Best Local Similarity 100.0%; Pred. No. 0; Matches 2041; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	TCCTTCACAGCATTTGGAGCTCTGAAGACAAAGTCAATTTGGCAGTTTCATTGATGGAGAG	60
Db	960	TCCTTCACAGCATTTGGAGCTCTGAAGACAAAGTCAATTTGGCAGTTTCATTGATGGAGAG	1019
QY	61	TGAAGTAAACCTGGACCGTTATCAACACAGCTTTAGAGAAAGTATATCGTGGCTTCTTTC	120
Db	1020	TGAAGTAAACCTGGACCGTTATCAACACAGCTTTAGAGAAAGTATATCGTGGCTTCTTTC	1079
QY	121	TGCTGAGGACACATTCGCAAGCACAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA	180
Db	1080	TGCTGAGGACACATTCGCAAGCACAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA	1139
QY	181	CCAGTTTCATCTACTGAGGGGTACATGATGATTTGACAGCCCCATCAGGCGCGGTGG	240
Db	1140	CCAGTTTCATCTACTGAGGGGTACATGATGATTTGACAGCCCCATCAGGCGCGGTGG	1199
QY	241	TAATATTTCTACAATTTGGGAAGTAAGCTGATGGAACAGGAAAAATATCAGAAGATGAAGA	300
Db	1200	TAATATTTCTACAATTTGGGAAGTAAGCTGATGGAACAGGAAAAATATCAGAAGATGAAGA	1259
QY	301	AACTGAAGTACAAGAGCAGATGAATCTCCATAATTCAGATGGGAATGCCCTCAGGCTAGC	360
Db	1260	AACTGAAGTACAAGAGCAGATGAATCTCCATAATTCAGATGGGAATGCCCTCAGGCTAGC	1319
QY	361	TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAAACT	420
Db	1320	TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAAACT	1379
QY	421	GAAAGAGTTGAATGACTGGCTAAACAAAAACAGAGAAAGAAACAAAGGAAATGGAGGAAGA	480
Db	1380	GAAAGAGTTGAATGACTGGCTAAACAAAAACAGAGAAAGAAACAAAGGAAATGGAGGAAGA	1439
QY	481	GCCTCTTGGACCTGATCTTGAAGACCTAAACCGCCCAAGTACAAACAAACATAAGGTGCTTCA	540
Db	1440	GCCTCTTGGACCTGATCTTGAAGACCTAAACCGCCCAAGTACAAACAAACATAAGGTGCTTCA	1499
QY	541	AGAAGATCTAGAACAGAACAGTCAAGGTCAATTTCTCCTCACTCACATGGTGGTGGTAGT	600
Db	1500	AGAAGATCTAGAACAGAACAGTCAAGGTCAATTTCTCCTCACTCACATGGTGGTGGTAGT	1559
QY	601	TGATGAATCTAGTGGAGATCAGGCAACTGCTGCTTTGGAGAAACAACTTAAGGTATTGGG	660
Db	1560	TGATGAATCTAGTGGAGATCAGGCAACTGCTGCTTTGGAGAAACAACTTAAGGTATTGGG	1619
QY	661	AGATCGATGGGCAACATCTGTAGATGGACAGAGACCGGTCTTTTACAAGACAG	720
Db	1620	AGATCGATGGGCAACATCTGTAGATGGACAGAGACCGGTCTTTTACAAGACAG	1679

QY	721	TTCTGACAGTGAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGTGGTACAGCT	780
Db	1680	TTCTGACAGTGAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGTGGTACAGCT	1739
QY	781	GAAAGATGATGAATTAAGCCGGCAGGCGACCTATTTGGAGGCGACTTCCAGCAGTTCAGAA	840
Db	1740	GAAAGATGATGAATTAAGCCGGCAGGCGACCTATTTGGAGGCGACTTCCAGCAGTTCAGAA	1799
QY	841	GCAGAACGATGTACATAGGGCCTTCAAGAGGGGAATTTGAAAGAACTTAAAGAACCTGTATCAT	900
Db	1800	GCAGAACGATGTACATAGGGCCTTCAAGAGGGGAATTTGAAAGAACTTAAAGAACCTGTATCAT	1859
QY	901	GAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGAAGGACTAGAGAA	960
Db	1860	GAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGAAGGACTAGAGAA	1919
QY	961	ACTTACCAGGAGGCCCGCAGAGCTGCCTCCTGAGGAGAGAGCCCGAGAAATGTCACTCGGCT	1020
Db	1920	ACTTACCAGGAGGCCCGCAGAGCTGCCTCCTGAGGAGAGAGCCCGAGAAATGTCACTCGGCT	1979
QY	1021	TCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTTGAACCTGCACCTCCG	1080
Db	1980	TCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTTGAACCTGCACCTCCG	2039
QY	1081	TGACTGGCAGAGAAAAATAGATGAGACCTTTGAAAGACTCCAGGAACCTTCAAGAGGCCAC	1140
Db	2040	TGACTGGCAGAGAAAAATAGATGAGACCTTTGAAAGACTCCAGGAACCTTCAAGAGGCCAC	2099
QY	1141	GGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGT	1200
Db	2100	GGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGT	2159
QY	1201	GGCGGATCTCCTCATTTGACTCTCTCCAAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGG	1260
Db	2160	GGCGGATCTCCTCATTTGACTCTCTCCAAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGG	2219
QY	1261	AGAAATTCGCGCTCTGAAAGAGAACGTCGAGCCACGTCATGACCTTGTCTGCCAGCTTAC	1320
Db	2220	AGAAATTCGCGCTCTGAAAGAGAACGTCGAGCCACGTCATGACCTTGTCTGCCAGCTTAC	2279
QY	1321	CACCTTGGGCATTCAGCTCTCACCGTATAAAGCTCAGCAGCTCTGGAAGACCTGAAACACAG	1380
Db	2280	CACCTTGGGCATTCAGCTCTCACCGTATAAAGCTCAGCAGCTCTGGAAGACCTGAAACACAG	2339
QY	1381	ATGGAAGCTTCTGCAGGTGGCGCTGAGGACCGAGTCAGGAGCTGCATGAAGCCACAG	1440
Db	2340	ATGGAAGCTTCTGCAGGTGGCGCTGAGGACCGAGTCAGGAGCTGCATGAAGCCACAG	2399
QY	1441	GGACTTTGGTCCAGCATCTCAGCAGCTTCTTCCACGCTCTGTCCAGGGTCCCTGGGAGAG	1500
Db	2400	GGACTTTGGTCCAGCATCTCAGCAGCTTCTTCCACGCTCTGTCCAGGGTCCCTGGGAGAG	2459
QY	1501	AGCCATCTCGCCAAACAAAGTCCCTACTATATATCAACACGAGACTCAAAACAACTTGCTG	1560
Db	2460	AGCCATCTCGCCAAACAAAGTCCCTACTATATCAACACGAGACTCAAAACAACTTGCTG	2519
QY	1561	GGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAATGTGATCT	1620
Db	2520	GGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAATGTGATCT	2579
QY	1621	CTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAGGCCCTTTGCTTGGATCT	1680
Db	2580	CTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAGGCCCTTTGCTTGGATCT	2639
QY	1681	CTTGAGCCTGTGAGCTGCATGTCCTTGGACCAGCAGCAACCTCAAGCAAAATGACCA	1740
Db	2640	CTTGAGCCTGTGAGCTGCATGTCCTTGGACCAGCAGCAACCTCAAGCAAAATGACCA	2699
QY	1741	GCCCATGGATATCCTGCAGATTAATTTGTTGACCACCTATTATGACCGCTGGAGCA	1800
Db	2700	GCCCATGGATATCCTGCAGATTAATTTGTTGACCACCTATTATGACCGCTGGAGCA	2759
QY	1801	AGAGCACAAACAATTTGGTCAACGTCCTCTCTGCGTGGATATGTGTGAACTGGCTGCT	1860

Db 2760 AGAGCACAACAATTGGTCAACGTCCTCTCTGCGTGGATATGIGTGAACCTGGCTGCT 2819
QY 1861 GAATGTTTATGATACGGGACGACAGGGAGGATCCGTGTCCTGTCTTTTAAAACTGGCAT 1920
Db 2820 GAATGTTTATGATACGGGACGACAGGGAGGATCCGTGTCCTGTCTTTTAAAACTGGCAT 2879
QY 1921 CATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAGATACCTTTTCAAGCAAGTGGC 1980
Db 2880 CATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAGATACCTTTTCAAGCAAGTGGC 2939
QY 1981 AAGTTCAACAGGATTTGTGACCGACGCGAGGCTGGGCTCCTTCTGTCATGATTTCTATCCA 2040
Db 2940 AAGTTCAACAGGATTTGTGACCGACGCGAGGCTGGGCTCCTTCTGTCATGATTTCTATCCA 2099
QY 2041 A 2041
Db 3000 A 3000

RESULT 2
AAD37259
ID AAD37259 standard; DNA; 4476 BP.
AC AAD37259;
XX
DT 21-AUG-2002 (first entry)
XX
DE Adeno-associated virus vector plasmid, AAV-MCK-3510.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Unidentified.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
XX
PI Xiao X;
XX
DR WPI; 2002-049342/06.
XX

PT New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -
XX
PS Example 1; Page 63-65; 71pp; English.
XX
CC The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression control
CC element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is AAV
CC vector plasmid construct containing human dystrophin minigenes, a muscle
CC creatine kinase (MCK) promoter and a small polyA signal sequence.
XX
SQ Sequence 4476 BP; 1252 A; 1096 C; 1127 G; 1001 T; 0 other;

Query Match 100.0%; Score 2041; DB 24; Length 4476;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2041; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TCCTTCACAGCATTTGGAGCTCCTGAAGACAAGTCAATTTGGCAGTTCATTGATGGAGAG 60
Db 1716 TCCTTCACAGCATTTGGAGCTCCTGAAGACAAGTCAATTTGGCAGTTCATTGATGGAGAG 1775
QY 61 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAGAACTATATCGTGGCTTCTTTC 120
Db 1776 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAGAACTATATCGTGGCTTCTTTC 1835
QY 121 TGCTGAGGACACATTGCAAGCACAAAGGAGAGATTTCTAATGATGTGGAAGTGGTAAAGA 180
Db 1836 TGCTGAGGACACATTGCAAGCACAAAGGAGAGATTTCTAATGATGTGGAAGTGGTAAAGA 1895
QY 181 CCAGTTTTCATCTATGAGGGGTACATGATGGATTTGACAGCCCATCAGGGCCGGTGG 240
Db 1896 CCAGTTTTCATCTATGAGGGGTACATGATGGATTTGACAGCCCATCAGGGCCGGTGG 1955
QY 241 TAATATTCTACAATTGGGAAGTAAGCTGATTGGACAGGAAAATATCAGAAGATGAAGA 300
Db 1956 TAATATTCTACAATTGGGAAGTAAGCTGATTGGACAGGAAAATATCAGAAGATGAAGA 2015
QY 301 AACTGAAGTACAAAGAGCAGATGAATCTCCTAAATCAAGATGGAATGCCTCAGGSTAGC 360
Db 2016 AACTGAAGTACAAAGAGCAGATGAATCTCCTAAATCAAGATGGAATGCCTCAGGSTAGC 2075
QY 361 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAAAT 420
Db 2076 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAAAT 2135
QY 421 GAAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGAAACAAAGAAAAATGGAGGAAGA 480
Db 2136 GAAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGAAACAAAGAAAAATGGAGGAAGA 2195
QY 481 GCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCAAGTACAACAACATAAGGTGCTTCA 540
Db 2196 GCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCAAGTACAACAACATAAGGTGCTTCA 2255
QY 541 AGAAGATCTAGAACAAAGAACACAGTCAAGGTCAATCTCTCACTACATGGTGGTAGT 600
Db 2256 AGAAGATCTAGAACAAAGAACACAGTCAAGGTCAATCTCTCACTACATGGTGGTAGT 2315
QY 601 TGATGAATCTAGTGGAGATCAGCAACTGCTGCTTTGGAAGAACTTAAGGTATTGGG 660
Db 2316 TGATGAATCTAGTGGAGATCAGCAACTGCTGCTTTGGAAGAACTTAAGGTATTGGG 2375
QY 661 AGATCGATGGGCAACATCTGTAGATGGACAGACCGCTGGTCTTTTACAAGACAG 720
Db 2376 AGATCGATGGGCAACATCTGTAGATGGACAGACCGCTGGTCTTTTACAAGACAG 2435
QY 721 TTCTGACCAGTGGAAAGCTGTCACCTTTCTCTGAGGAACCTTCTGGTGTGGCTACAGCT 780
Db 2436 TTCTGACCAGTGGAAAGCTGTCACCTTTCTCTGAGGAACCTTCTGGTGTGGCTACAGCT 2495
QY 781 GAAAGATGATGAATTAAGCCGGCAGGCACCTATTGGAGGCGGACTTTCCAGCAGTTCAGAA 840
Db 2496 GAAAGATGATGAATTAAGCCGGCAGGCACCTATTGGAGGCGGACTTTCCAGCAGTTCAGAA 2555
QY 841 GCAGAACGATGTACATAGGCGCTTCAAGAGGGAATTTGAAACTAAAGAACCTGTAAATCAT 900
Db 2556 GCAGAACGATGTACATAGGCGCTTCAAGAGGGAATTTGAAACTAAAGAACCTGTAAATCAT 2615
QY 901 GAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAAGGACTAGAGAA 960
Db 2616 GAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAAGGACTAGAGAA 2675
QY 961 ACTCTACCAGGAGCCAGAGAGCTGCCTCCTGAGGAGAGAGCCCGAGAAATGTCACTCGGCT 1020
Db 2676 ACTCTACCAGGAGCCAGAGAGCTGCCTCCTGAGGAGAGAGCCCGAGAAATGTCACTCGGCT 2735
QY 1021 TCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGAAAAATTTGAACCTGCACCTCCGC 1080

Db 2736 TCTACGAAGCAGGCTGAGGAGTCAATACTAGTGGGAAAAATTTGAACCTGCACCTCCGC 2795
QY 1081 TGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAAGACTCCAGGAATTCAGAGGCCAC 1140
Db 2796 TGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAAGACTCCAGGAATTCAGAGGCCAC 2855
QY 1141 GGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGT 1200
Db 2856 GGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGT 2915
QY 1201 GGGCGATCTCCTCATTTGACTCTCTCCAAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGG 1260
Db 2916 GGGCGATCTCCTCATTTGACTCTCTCCAAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGG 2975
QY 1261 AGAAATTTGGCCTCTGAAAGAGAACGCTGAGCCACGTTCAATGACCTTGCTCGCCAGCTTAC 1320
Db 2976 AGAAATTTGGCCTCTGAAAGAGAACGCTGAGCCACGTTCAATGACCTTGCTCGCCAGCTTAC 3035
QY 1321 CACTTTGGGCAATTCAGCTCTCACCCTGATTAACCTCAGCACTCTGGAAGACCTGAACACCCAG 1380
Db 3036 CACTTTGGGCAATTCAGCTCTCACCCTGATTAACCTCAGCACTCTGGAAGACCTGAACACCCAG 3095
QY 1381 ATGGAAGCTTCTGAGGTGGCCGTCGAGGACCGAGTCAGCAGCTGCATGAAGCCACACAG 1440
Db 3096 ATGGAAGCTTCTGAGGTGGCCGTCGAGGACCGAGTCAGCAGCTGCATGAAGCCACACAG 3155
QY 1441 GGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGCTCTGCCAGGGTCCCTGGGAGAG 1500
Db 3156 GGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGCTCTGCCAGGGTCCCTGGGAGAG 3215
QY 1501 AGCCATCTCGCCAAACAAAGTGGCCCTACTATATCAACCACGAGACTCAAAACAACTTGCTG 1560
Db 3216 AGCCATCTCGCCAAACAAAGTGGCCCTACTATATCAACCACGAGACTCAAAACAACTTGCTG 3275
QY 1561 GGACCATCCCAAAATGACAGAGCTCTACCACTCTTTAGCTGACCTGAATATGTCAGATT 1620
Db 3276 GGACCATCCCAAAATGACAGAGCTCTACCACTCTTTAGCTGACCTGAATATGTCAGATT 3335
QY 1621 CTCAGCTTATAGGACTGCCATGAACTCCGAAGACTGCAGAGGGCCCTTTGCTTGGATCT 1680
Db 3336 CTCAGCTTATAGGACTGCCATGAACTCCGAAGACTGCAGAGGGCCCTTTGCTTGGATCT 3395
QY 1681 CTTGAGCTGTGACGTGCATGTGATGCCCTGGACCAAGCAACCTCAAGCAAAATGACCA 1740
Db 3396 CTTGAGCTGTGACGTGCATGTGATGCCCTGGACCAAGCAACCTCAAGCAAAATGACCA 3455
QY 1741 GCCCATGGATPCTCTGCAGATTATTAATTGTTGACCACTATTTATGACCCGCTGGAGCA 1800
Db 3456 GCCCATGGATPCTCTGCAGATTATTAATTGTTGACCACTATTTATGACCCGCTGGAGCA 3515
QY 1801 AGAGCACAAACAAATTTGGTCAACGTCCTCTCTGCGTGGATATGTGCTGAACTGGCTGCT 1860
Db 3516 AGAGCACAAACAAATTTGGTCAACGTCCTCTCTGCGTGGATATGTGCTGAACTGGCTGCT 3575
QY 1861 GAATGTTTATGATACGGGACGAACAGGGAGGATCCCTGCTGCTCTTTTAAACTGGCAT 1920
Db 3576 GAATGTTTATGATACGGGACGAACAGGGAGGATCCCTGCTGCTCTTTTAAACTGGCAT 3635
QY 1921 CATTTCCTGTGTAAAGCACATTTGGAAGACAAGTACAGATACCTTTTCAAGCAAGTGGC 1980
Db 3636 CATTTCCTGTGTAAAGCACATTTGGAAGACAAGTACAGATACCTTTTCAAGCAAGTGGC 3695
QY 1981 AAGTTCAACAGGATTTTGTGACCAAGCGCAGGCTGGGCTCCTCTCTGTCATGATCTATCCA 2040
Db 3696 AAGTTCAACAGGATTTTGTGACCAAGCGCAGGCTGGGCTCCTCTCTGTCATGATCTATCCA 3755

AAD37238
ID AAD37238 standard; DNA; 3531 BP.
XX
AC AAD37238;
XX
DT 21-AUG-2002 (first entry)
XX
DE Human dystrophin minigene delta3531.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Homo sapiens.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
XX
PI Xiao X;
XX
DR WPI; 2002-049342/06.
XX
XX New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -
XX
PS Example 1; Page 50-51; 71pp; English.

The present invention relates to an isolated nucleotide sequence encoding a dystrophin minigene. The minigene comprises N-terminal or modified N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control element. The dystrophin minigene in operable linkage with an expression control element, in a recombinant adeno-associated virus or retrovirus is useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is human dystrophin minigene delta3531 containing nucleotides 1-1341 (N-terminus, hinge H1 and rod R1), 8059-10227 (rods R22, R23 and R24, hinge H4 and CR domain) and 11047-11058 (dystrophin last 3 amino acids).

Sequence 3531 BP; 1071 A; 809 C; 824 G; 827 T; 0 other;

Query Match 82.8%; Score 1690.8; DB 24; Length 3531;
Best Local Similarity 89.9%; Pred. No. 0;
Matches 1854; Conservative 0; Mismatches 187; Indels 21; Gaps 3;

QY 1 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAAGTCAATTTGGCAGTTCATTGATGGAGAG 60
Db 960 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAAGTCAATTTGGCAGTTCATTGATGGAGAG 1019
QY 61 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAAATATATCGTGGCTTCTTTC 120
Db 1020 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAAATATATCGTGGCTTCTTTC 1079
QY 121 TGCTGAGGACACATTGCAAGCACAAAGGAGAGATTTCTAATGATGTGGAGTGGTGAAGA 180
Db 1080 TGCTGAGGACACATTGCAAGCACAAAGGAGAGATTTCTAATGATGTGGAGTGGTGAAGA 1139
QY 181 CCAGTTTCATACCTCATGAGGGGTACATGATGGATTTGACAGCCCATCAGGGCCGGTGG 240
Db 1140 CCAGTTTCATACCTCATGAGGGGTACATGATGGATTTGACAGCCCATCAGGGCCGGTGG 1199
QY 241 TAATATTCTACAATTGGGAAGTAAGCTGATTGGAAACAGGAAAAATTATCAGAAGATGAAGA 300

Db 1200 TAAATATTCTACAATTTGGGAAGTAAGCTGATTGGAAACAGGAAAAATATCAGAAGATGAAGA 1259
QY 301 AACTGAAGTACAAGAGCAGATGAATCTCCTAATAATTCAAGATGGGAATGCCTCAGGGTAGC 360
Db 1260 AACTGAAGTACAAGAGCAGATGAATCTCCTAATAATTCAAGATGGGAATGCCTCAGGGTAGC 1319
QY 361 TAGCATGGAAGAAACAAAGCAATTTACATAGAGTT-----TTAATGGATCTCCAGAATCA 414
Db 1320 TAGCATGGAAGAAACAAAGCAATTTACATAGAACTCATAGATTACTGCAACAGTTCCTCCCT 1379
QY 415 GAAACTGAAAGAGTTGAATGACTGGCTTAACAAAAACAGAAAGAAAGCAAGCAAGGAAATGGA 474
Db 1380 GGACCTGGAAGAGTTTCTTGCCCTGGCTTACAGAAAGCTGAAACAACTGCCAATGTCCTACA 1439
QY 475 GGAAGAGCCCTCTTGGACCTGATCTTGAAGACCTTAAAGCGCAAGTACAAACACATAAAGGT 534
Db 1440 GGATGCTACCCGTAAGGAAAGGCTCCTAGAAGACTCCAAGGGAGTAAAAGAGCTGATGAA 1499
QY 535 GCTTCAAGAAGATCTAGAACAAAGAAACAAAGTCAAGGTCAATTTCTCAGCTCAGATGGTGGT 594
Db 1500 ACAATGGCAAGACCTCCAGAGTGAATTTGAAGCTCACACAGATGTTTATCACAACTGGA 1559
QY 595 GGTAGTTGATGA-----ATCTAGTGGAGATCACGCAACTGCTGCTCTTTGGAAGA 642
Db 1560 TGAANAACAGCCAAAATACTCTGAGATCCCTGGAAGGTTCCGATGATGAGTCTCTGTACA 1619
QY 643 ACAACTTAAGG---TATTGGGAGATCGATGGCAAAACATCTGTAGATGGACAGAACCG 699
Db 1620 AAGACGTTTGGATAACATGAACCTTCAAGTGGAGTGAACCTCGGAAAAAGTCTCTCAACAT 1679
QY 700 CTGGGTTCTTTTACAAGACAGTTCTGACCAGTGGAAAGCGTCTGACCTTTCTCTGCAGGA 759
Db 1680 TAGGTCCCATTTGGAAGCCAGTTCTGACCAGTGGAAAGCGTCTGACCTTTCTCTGCAGGA 1739
QY 760 ACTTCTGGTGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTGGAGG 819
Db 1740 ACTTCTGGTGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTGGAGG 1799
QY 820 CGACTTTCCAGCAGTTTCAAGAGCAGAACGATGTACATAGGGCCTTCAAGAGGGAATTGAA 879
Db 1800 CGACTTTCCAGCAGTTTCAAGAGCAGAACGATGTACATAGGGCCTTCAAGAGGGAATTGAA 1859
QY 880 AACTAAAGAACCTGTAAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCA 939
Db 1860 AACTAAAGAACCTGTAAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCA 1919
QY 940 GCCTTTGGAAGGACTAGAGAAACTCTACCAGGAGCCCAAGAGCTGCCTCCTGAGGAGAG 999
Db 1920 GCCTTTGGAAGGACTAGAGAAACTCTACCAGGAGCCCAAGAGCTGCCTCCTGAGGAGAG 1979
QY 1000 AGCCCAAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGA 1059
Db 1980 AGCCCAAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGA 2039
QY 1060 AAAATTTGAACCTGCACCTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGAGACT 1119
Db 2040 AAAATTTGAACCTGCACCTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGAGACT 2099
QY 1120 CCAGGAACCTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTCGCCCAAGCTGAGGTGAT 1179
Db 2100 CCAGGAACCTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTCGCCCAAGCTGAGGTGAT 2159
QY 1180 CAAGGGATCCTGGCAGCCCGTGGCGGATCTCCTCATTTGACTCTCTCCAAAGATCACCTCGA 1239
Db 2160 CAAGGGATCCTGGCAGCCCGTGGCGGATCTCCTCATTTGACTCTCTCCAAAGATCACCTCGA 2219
QY 1240 GAAAGTCAAGGCACTTCGAGGAGAAAAATTGCGCCCTCTGAAAGAGAGACGTGAGCCACGTCAA 1299
Db 2220 GAAAGTCAAGGCACTTCGAGGAGAAAAATTGCGCCCTCTGAAAGAGAGACGTGAGCCACGTCAA 2279
QY 1300 TGACCTTGTCTGCGCAGCTTACCACCTTTTGGGCATTTACCGCTCTACCGGTATAACCTCAGCAC 1359
Db 1300 TGAACCTTGTCTGCGCAGCTTACCACCTTTTGGGCATTTACCGCTCTACCGGTATAACCTCAGCAC 1359

Db 2280 TGACCTTGTCTGCGCAGCTTACCACCTTTGGGCATTTACGCTCTCACCGTATAAACCTCAGCAC 2339
QY 1360 TCTGGAAGACCTGAACACACAGATGGAAGCTTCTGAGGTGGCCGTCGAGGACCCGAGTCAG 1419
Db 2340 TCTGGAAGACCTGAACACACAGATGGAAGCTTCTGAGGTGGCCGTCGAGGACCCGAGTCAG 2399
QY 1420 GCAGCTGCATGAAGCCCAACAGGAGCTTTGGTCCAGCATCTCAGCAGCTTTCTTTCCACGTC 1479
Db 2400 GCAGCTGCATGAAGCCCAACAGGAGCTTTGGTCCAGCATCTCAGCAGCTTTCTTTCCACGTC 2459
QY 1480 TGTCCAGGCTCCCTGGGAGAGAGGCCATCTCGCCAAAACAAAGTGCCCTACTATATCAACCA 1539
Db 2460 TGTCCAGGCTCCCTGGGAGAGAGGCCATCTCGCCAAAACAAAGTGCCCTACTATATCAACCA 2519
QY 1540 CGAGACTCAAACAACTTGTGGGAGAGGCCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGC 1599
Db 2520 CGAGACTCAAACAACTTGTGGGAGAGGCCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGC 2579
QY 1600 TGACCTGAATAATGTGAGATCTCAGCTTATAGGACTGCCATGAAACTCCGAAAGACTGCA 1659
Db 2580 TGACCTGAATAATGTGAGATCTCAGCTTATAGGACTGCCATGAAACTCCGAAAGACTGCA 2639
QY 1660 GAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTGAGCTGTCATGTCATGATGCCTTGGACCAGCA 1719
Db 2640 GAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTGAGCTGTCATGTCATGATGCCTTGGACCAGCA 2699
QY 1720 CAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTTGACCAC 1779
Db 2700 CAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTTGACCAC 2759
QY 1780 TATTTATGACCGCCTGGAGCAAGAGCACAACAATTTGGTCAACGTCCTCTCTGCGTGA 1839
Db 2760 TATTTATGACCGCCTGGAGCAAGAGCACAACAATTTGGTCAACGTCCTCTCTGCGTGA 2819
QY 1840 TATGTCGTGAACCTGGCTGCTGAATGTTTATGATACGGGACCAACAGGAGGATCCGTGT 1899
Db 2820 TATGTCGTGAACCTGGCTGCTGAATGTTTATGATACGGGACCAACAGGAGGATCCGTGT 2879
QY 1900 CCTGTCTTTAAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAG 1959
Db 2880 CCTGTCTTTAAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAG 2939
QY 1960 ATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTGTGACCAGCGCAGGTGGGCT 2019
Db 2940 ATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTGTGACCAGCGCAGGTGGGCT 2999
QY 2020 CCTTCTGCATGATTTCTATCCAA 2041
Db 3000 CCTTCTGCATGATTTCTATCCAA 3021

RESULT 4
AAD37258
ID AAD37258 standard; DNA; 4498 BP.
XX
AC AAD37258;
XX
DT 21-AUG-2002 (first entry)
XX
DE Adeno-associated virus vector plasmid, AAV-MCK-3531.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
XX Becker muscular dystrophy; ds.
OS Chimeric - Homo sapiens.
OS Chimeric - Unidentified.
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.

28-APR-2000; 2000US-200777P.

(XIAO/) XIAO X.

Xiao X;

WPI; 2002-049342/06.

New dystrophin minigene for treating Duchenne or Becker muscular dystrophy comprises an N-terminal domain or modified N-terminal domain, rod repeats, H1 and H4 domains and a cysteine rich domain of a dystrophin gene.

Example 1; Page 62-63; 71pp; English.

The present invention relates to an isolated nucleotide sequence encoding a dystrophin minigene. The minigene comprises N-terminal or modified N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control element. The dystrophin minigene in operable linkage with an expression control element, in a recombinant adeno-associated virus or retrovirus is useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is AAV vector plasmid construct containing human dystrophin minigenes, a muscle creatine kinase (MCK) promoter and a small polyA signal sequence.

Sequence 4498 BP; 1251 A; 1118 C; 1123 G; 1006 T; 0 other;

Query Match 82.8%; Score 1690.8; DB 24; Length 4498;

Best Local Similarity 89.9%; Pred. No. 0;

Matches 1854; Conservative 0; Mismatches 187; Indels 21; Gaps 3;

QY	1	TCCTTACAGCATTTGGAGCTCCTGAAGACAAAGTCATTTGGCAGTTTCATTGATGGAGAG	60
DB	1717	TCCTTACAGCATTTGGAGCTCCTGAAGACAAAGTCATTTGGCAGTTTCATTGATGGAGAG	1776
QY	61	TGAAGTAAACCTGGACCGTTATCAACACAGCTTTAGAAGAAATATATCGTGGCTTCTTTC	120
DB	1777	TGAAGTAAACCTGGACCGTTATCAACACAGCTTTAGAAGAAATATATCGTGGCTTCTTTC	1836
QY	121	TGCTGAGGACACATTTGCAAGCACAAAGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA	180
DB	1837	TGCTGAGGACACATTTGCAAGCACAAAGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA	1896
QY	181	CCAGTTTCATACATCATGAGGGGTACATGATGATGATGATGATGATGATGATGATGATGAT	240
DB	1897	CCAGTTTCATACATCATGAGGGGTACATGATGATGATGATGATGATGATGATGATGATGAT	1956
QY	241	TAATATTCTACAATTGGGAAGTAAGCTGATGATGATGATGATGATGATGATGATGATGAT	300
DB	1957	TAATATTCTACAATTGGGAAGTAAGCTGATGATGATGATGATGATGATGATGATGATGAT	2016
QY	301	AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGATGGGAATGCCTCAGGGTAGC	360
DB	2017	AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGATGGGAATGCCTCAGGGTAGC	2076
QY	361	TAGCATGGAAGAAACAAAGCAATTTACATAGAGTT-----TTAATGGATCTCCAGAAATCA	414
DB	2077	TAGCATGGAAGAAACAAAGCAATTTACATAGAGTTACTGCAACAGTTTCCCTTCCCTT	2136
QY	415	GAAAGTGAAGAGTTGAATGACTGGCTTAACAAACAGAGAGAGAGAGAGAGAGAGAGAGAG	474
DB	2137	GGACCTGGAAAGTTTCTTGCCTGGCTTACAGAGAGTGAACAACTGCCAATGTCTCTACA	2196
QY	475	GGAAGAGCCTCTTGGACCTGATCTTGAAGACCTAAACCGCCCAAGTACAACAACATAAGGT	534
DB	2197	GGATGCTACCCGTAAGGAAAGGCTCCTAGAAGACTCCAGGGAGTAAAGAGCTGATGAA	2256
QY	535	GCTTCAAGAAGATCTAGAACAAGAAAGTCAAGGTCAATTTCTCTCACTCACATGGTGGT	594

QY 1660 GAAGGCCCTTTGCTGGATCTCTTGAGCCTGTGAGCCTGCATGTGATGCCTTGGACCAGCA 1719
Db 3397 GAAGGCCCTTTGCTGGATCTCTTGAGCCTGTGAGCCTGCATGTGATGCCTTGGACCAGCA 3456
QY 1720 CAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTTGACCAC 1779
Db 3457 CAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTTGACCAC 3516
QY 1780 TATTTATGACCGCCTGGAGCAAGAGCACAAATTTGGTCAACGTCCCTCTCTGCGTGA 1839
Db 3517 TATTTATGACCGCCTGGAGCAAGAGCACAAATTTGGTCAACGTCCCTCTCTGCGTGA 3576
QY 1840 TATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCCGTGT 1899
Db 3577 TATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCCGTGT 3636
QY 1900 CCTGTCTTTTAAAACTGGCATCATTTCCCTGTGTAAAGCACACATTTGGAAGACAAAGTACAG 1959
Db 3637 CCTGTCTTTTAAAACTGGCATCATTTCCCTGTGTAAAGCACACATTTGGAAGACAAAGTACAG 3696
QY 1960 ATACCTTTTCAAGCAAGTGGCAAGTTCACAGGATTTTGTGACCGAGCGAGGCTGGSCCT 2019
Db 3697 ATACCTTTTCAAGCAAGTGGCAAGTTCACAGGATTTTGTGACCGAGCGAGGCTGGSCCT 2078
QY 2020 CCTTCTGCATGATTTCTATCCAA 2041
Db 3757 CCTTCTGCATGATTTCTATCCAA 3778

RESULT 5

AAD37237
ID AAD37237 standard; DNA; 3858 BP.
XX AC AAD37237;
XX DT 21-AUG-2002 (first entry)
XX DE Human dystrophin minigene delta3849.
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX OS Homo sapiens.
XX PN WO200183695-A2.
XX PD 08-NOV-2001.
XX PP 27-APR-2001; 2001WO-US13677.
XX PR 28-APR-2000; 2000US-200777P.
PA (XIAO/) XIAO X.
XX XIAO X;
XX WPI; 2002-049342/06.
XX

New dystrophin minigene for treating Duchenne or Becker muscular dystrophy comprises an N-terminal domain or modified N-terminal domain, rod repeats, H1 and H4 domains and a cysteine rich domain of a dystrophin gene -

Example 1; Page 48-49; 71pp; English.

The present invention relates to an isolated nucleotide sequence encoding a dystrophin minigene. The minigene comprises N-terminal or modified N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control

CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is human
CC dystrophin minigene delta3849 containing nucleotides 1-1668 (N-terminus,
CC hinge H1 and rods R1, R2), 8059-10227 (rods R22, R23 and R24, hinge H4
CC and CR domain) and 11047-11058 (dystrophin last 3 amino acids).
XX

SQ Sequence 3858 BP; 1189 A; 866 C; 905 G; 898 T; 0 other;

Query Match 82.5%; Score 1683; DB 24; Length 3858;
Best Local Similarity 85.4%; Pred. NO. 0;
Matches 2041; Conservative 0; Mismatches 0; Indels 348; Gaps 1;

QY 1 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAAGTCAATTTGGCAGTTCATTGATGGAGAG 60
Db 960 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAAGTCAATTTGGCAGTTCATTGATGGAGAG 1019
QY 61 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGATATTATCGTGGCTTCTTTC 120
Db 1020 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGATATTATCGTGGCTTCTTTC 1079
QY 121 TGCTGAGGACACATTGCAAGCACAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 180
Db 1080 TGCTGAGGACACATTGCAAGCACAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 1139
QY 181 CCAGTTTCACTCATGAGGGGTACATGATGGATTGACAGCCCATCAGGGCCGGTGG 240
Db 1140 CCAGTTTCACTCATGAGGGGTACATGATGGATTGACAGCCCATCAGGGCCGGTGG 1199
QY 241 TAATATTCTACAATTGGGAAGTAAAGCTGATTGGAACAGGAAATTTATCAGAAGATGAAGA 300
Db 1200 TAATATTCTACAATTGGGAAGTAAAGCTGATTGGAACAGGAAATTTATCAGAAGATGAAGA 1259
QY 301 AACTGAAGTACAAGACAGATGAATCTCCTAAATTCAGATGGGATGCCTCAGGGTAGC 360
Db 1260 AACTGAAGTACAAGACAGATGAATCTCCTAAATTCAGATGGGATGCCTCAGGGTAGC 1319
QY 361 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTTTAATGGATCTCCAGAATCAGAAACT 420
Db 1320 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTTTAATGGATCTCCAGAATCAGAAACT 1379
QY 421 GAAAGAGTTGAATGACTGGCTAACAAAAACAGAGAAAGAAACAAAGGAAAAATGGAGGAAGA 480
Db 1380 GAAAGAGTTGAATGACTGGCTAACAAAAACAGAGAAAGAAACAAAGGAAAAATGGAGGAAGA 1439
QY 481 GCCTCTTGGACCTGATCTTGAAGACCTAAACCGCAAGTACAACAAACATAAGGTGCTTCA 540
Db 1440 GCCTCTTGGACCTGATCTTGAAGACCTAAACCGCAAGTACAACAAACATAAGGTGCTTCA 1499
QY 541 AGAAGATCTAGAACAAAGAACAAAGTCAAGGTCAATTTCTCACTCACATGGTGGTAGT 600
Db 1500 AGAAGATCTAGAACAAAGAACAAAGTCAAGGTCAATTTCTCACTCACATGGTGGTAGT 1559
QY 601 TGATGAATCTAGTGGAGATCACCGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGG 660
Db 1560 TGATGAATCTAGTGGAGATCACCGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGG 1619
QY 661 AGATCGATGGGCAACATCTGTAGATGGACAGAACCGCTGGGTCTCTTTTACAAGA --- 717
Db 1620 AGATCGATGGGCAACATCTGTAGATGGACAGAACCGCTGGGTCTCTTTTACAAGACAC 1679
QY 718 ----- 717
Db 1680 TCATAGATTACTGCAACAGTTCCCTCCCTGGACCTGGAAAAAGTTTCTGCTGGCTTACAGA 1739
QY 718 ----- 717
Db 1740 AGCTGAACAACACTGCCAATGTCTACAGGATGCTACCCCGTAAGGAAAGGCTCCTAGAAGA 1799
QY 718 ----- 717
Db 1800 CTCCAAGGGAGTAAAAAGAGCTGATGAAACAATGGCAACACCTCCAAGGTGAAATTTGAAGC 1859

QY	1633	GACTGCCATGAAACTCCGAAGACTGCAGAAGCCCTTTGCTTGGATCTCTTGAGCCTGTC	1692
Db	2940	GACTGCCATGAAACTCCGAAGACTGCAGAAGCCCTTTGCTTGGATCTCTTGAGCCTGTC	2999
QY	1693	AGCTGCATGTGATGCCCTTGGACCAACCTCAAGCAAAATGACCAGCCCATGGATAT	1752
Db	3000	AGCTGCATGTGATGCCCTTGGACCAACCTCAAGCAAAATGACCAGCCCATGGATAT	3059
QY	1753	CCTGCAGATTATTAATTGTTTGACCACCTATTATGACCGCCTGGAGCAAGAGCACAACAA	1812
Db	3060	CCTGCAGATTATTAATTGTTTGACCACCTATTATGACCGCCTGGAGCAAGAGCACAACAA	3119
QY	1813	TTTGGTCAACGTCCCTCTCTGCGTGGATATGTGTCTGAACCTGGCTGCTGAATGTTTATGA	1872
Db	3120	TTTGGTCAACGTCCCTCTCTGCGTGGATATGTGTCTGAACCTGGCTGCTGAATGTTTATGA	3179
QY	1873	TACGGGACGAACAGGGAGGATCCGTGTCTCTTTTAAAACTGGCATCATTTCCCTGTG	1932
Db	3180	TACGGGACGAACAGGGAGGATCCGTGTCTCTTTTAAAACTGGCATCATTTCCCTGTG	3239
QY	1933	TAAAGCACATTTGGAAGACAAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGG	1992
Db	3240	TAAAGCACATTTGGAAGACAAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGG	3299
QY	1993	ATTTTGTGACCAGCGAGCGTGGCCCTCCTTCTGCGATGATTCTATCCAA	2041
Db	3300	ATTTTGTGACCAGCGAGCGTGGCCCTCCTTCTGCGATGATTCTATCCAA	3348

RESULT 6

RESOLUTION 0
AAD37257
ID AAD37257 standard: DNA: 4825 BP.

AA
AC

XX
DT 21-AUG-2002 (first entry)

XX associated virus vector plasmid, AAV-MCK-delta3849.

XX Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.

XX	Chimeric	- Homo sapiens.
OS	Chimeric <td>- Identified.</td>	- Identified.
OS	Chimeric <td>- Identified.</td>	- Identified.

XX
PN
WO200183695-A2.XX
PD
08-NOV-2001.XX
PF 27-APR-2001: 2001WO-US13677.XX
PP 28-APR-2000: 2000US-200777P.XX (XIAO/) XIAO X.
PAXX
PT Xiao X;

AA
DR WPT: 2002-049342/06.

XX New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -
PT

XX
ps Example 1: Page 61-62; 71pp; English.

xx The present invention relates to an isolated nucleotide sequence encoding
cc a dystrophin minigene. The minigene comprises N-terminal or modified
cc N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
cc domains and cysteine-rich domains of dystrophin or utrophin genes. The
cc invention also relates to a recombinant adeno-associated virus (AAV)
cc comprising dystrophin minigene operably linked to an expression control

QY	718	-----	717
Db	1860	TCACACAGATGTTATCACAACTGGATGAAAACAGCCAAAAATCCTGAGATCCCTGGA	1919
QY	718	-----	717
Db	1920	AGGTTCCGATGATGCAGTCCCTGTTACAAAGACGTTTGGATAACATGAACCTTCAAGTGGAG	1979
QY	718	-----	732
Db	1980	TGAACCTTCGGAAAAAGTCTCTCAACATTAGGTCCCATTTGGAAGCCAGTTCTGACCAGTG	2039
QY	733	GAAGCGTCTGCACCTTTCTCTGCAGGAACTTCTGGTGGCTACAGCTGAAAGATGATGA	792
Db	2040	GAAGCGTCTGCACCTTTCTCTGCAGGAACTTCTGGTGGCTACAGCTGAAAGATGATGA	2099
QY	793	ATTAAGCCGGCAGGCACCTATTTGGAGGCGACTTTCAGCGAGTTCAGAAGCAGAACGATGT	852
Db	2100	ATTAAGCCGGCAGGCACCTATTTGGAGGCGACTTTCAGCGAGTTCAGAAGCAGAACGATGT	2159
QY	853	ACATAGGGCCTTCAAGAGGGGAATTTGAAAAACTAAAGAACCTGTAAATCATGAGTACTCTTGA	912
Db	2160	ACATAGGGCCTTCAAGAGGGGAATTTGAAAAACTAAAGAACCTGTAAATCATGAGTACTCTTGA	2219
QY	913	GACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAAGGACTAGAGAACTCTACCAGGA	972
Db	2220	GACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAAGGACTAGAGAACTCTACCAGGA	2279
QY	973	GCCCAGAGAGCTGCCTCCTGAGGAGAGAGCCAGAAATGTCACTCGGCTTCTACGAAAGCA	1032
Db	2280	GCCCAGAGAGCTGCCTCCTGAGGAGAGAGCCAGAAATGTCACTCGGCTTCTACGAAAGCA	2339
QY	1033	GGCTGAGGAGGTCAATACTAGTGGGAAAAATTTGAACCTGCACTCCGCTGACTGGCAGAG	1092
Db	2340	GGCTGAGGAGGTCAATACTAGTGGGAAAAATTTGAACCTGCACTCCGCTGACTGGCAGAG	2399
QY	1093	AAAAATAGATGAGACCCCTTGAAGACTCCAGGAACTTCAAGAGGCCACGGATGAGCTGGA	1152
Db	2400	AAAAATAGATGAGACCCCTTGAAGACTCCAGGAACTTCAAGAGGCCACGGATGAGCTGGA	2459
QY	1153	CCTCAAGCTGGCCAAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGGCGATCTCCT	1212
Db	2460	CCTCAAGCTGGCCAAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGGCGATCTCCT	2519
QY	1213	CATTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGCACTTCGAGGAGAAATTGCGCC	1272
Db	2520	CATTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGCACTTCGAGGAGAAATTGCGCC	2579
QY	1273	TCTGAAAGAGAACGTGAGCCACGTCAATGACCTTGCTCGCCAGCTTACCACCTTTGGGCAT	1332
Db	2580	TCTGAAAGAGAACGTGAGCCACGTCAATGACCTTGCTCGCCAGCTTACCACCTTTGGGCAT	2639
QY	1333	TCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCTGAACACCAGATGGAAGCTTCT	1392
Db	2640	TCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCTGAACACCAGATGGAAGCTTCT	2699
QY	1393	GCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGAAGCCCAACAGGACTTTGGTCC	1452
Db	2700	GCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGAAGCCCAACAGGACTTTGGTCC	2759
QY	1453	AGCATCTCAGCACTTTCTTTCCACGTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCC	1512
Db	2760	AGCATCTCAGCACTTTCTTTCCACGTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCC	2819
QY	1513	AAACAAAGTGCCCTACTATATCAACCACGAGACTCAAACAACCTTGCTGGGACCATCCCAA	1572
Db	2820	AAACAAAGTGCCCTACTATATCAACCACGAGACTCAAACAACCTTGCTGGGACCATCCCAA	2879
QY	1573	AATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAATGTGAGATTCTCAGCTTATAG	1632
Db	2880	AATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAATGTGAGATTCTCAGCTTATAG	2939

CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is AAV
CC vector plasmid construct containing human dystrophin minigenes, a muscle
CC creatine kinase (MCK) promoter and a small polyA signal sequence.
XX
SQ Sequence 4825 BP; 1369 A; 1175 C; 1204 G; 1077 T; 0 other;

Query Match 82.5%; Score 1683; DB 24; Length 4825;
Best Local Similarity 85.4%; Pred. No. 0;
Matches 2041; Conservative 0; Mismatches 0; Indels 348; Gaps 1;

QY 1 TCCTTCACAGCATTTGGAAAGCTCCTGAAGACAAAGTCAATTTGGCAGTTTCATTGATGGAGAG 60
Db 1717 TCCTTCACAGCATTTGGAAAGCTCCTGAAGACAAAGTCAATTTGGCAGTTTCATTGATGGAGAG 1776
QY 61 TGAAGTAAACCTGGACCGTTATCAACAGCTTTAGAAAGAGTATTATCGTGGCTTCTTTC 120
Db 1777 TGAAGTAAACCTGGACCGTTATCAACAGCTTTAGAAAGAGTATTATCGTGGCTTCTTTC 1836
QY 121 TGCTGAGGACACATTTGCAAGCACAAAGGAGAGATTTCTAATGATGTGGAAGTGTGAAAGA 180
Db 1837 TGCTGAGGACACATTTGCAAGCACAAAGGAGAGATTTCTAATGATGTGGAAGTGTGAAAGA 1896
QY 181 CCAGTTTCATACTCATGAGGGGTACATGATGGATTTGACAGCCCCATCAGGSCCGGTTGG 240
Db 1897 CCAGTTTCATACTCATGAGGGGTACATGATGGATTTGACAGCCCCATCAGGSCCGGTTGG 1956
QY 241 TAATATTCTACAAATTTGGGAAGTAAGCTGATTTGGAACAGGAAATTTATCAGAAGATGAAGA 300
Db 1957 TAATATTCTACAAATTTGGGAAGTAAGCTGATTTGGAACAGGAAATTTATCAGAAGATGAAGA 2016
QY 301 AACTGAAGTACAAAGACAGATGAATCTCCTAAATTTCAAGATGGAATGCCTCAGGTTAGC 360
Db 2017 AACTGAAGTACAAAGACAGATGAATCTCCTAAATTTCAAGATGGAATGCCTCAGGTTAGC 2076
QY 361 TAGCATGGAATAAACAAGCAATTTACATAGAGTTTTTAATGATCTCCAGAAATCAGAAACT 420
Db 2077 TAGCATGGAATAAACAAGCAATTTACATAGAGTTTTTAATGATCTCCAGAAATCAGAAACT 2136
QY 421 GAAAGAGTTGAATGACTGGCTAACAACAAACAGAGAAAGAAACAAAGAAATGGAGGAAGA 480
Db 2137 GAAAGAGTTGAATGACTGGCTAACAACAAACAGAGAAAGAAACAAAGAAATGGAGGAAGA 2196
QY 481 GCCTCTTGGACCTGATCTTTGAAGACCTTAAACCCCAAGTACAAACATAAGGTGCTTCA 540
Db 2197 GCCTCTTGGACCTGATCTTTGAAGACCTTAAACCCCAAGTACAAACATAAGGTGCTTCA 2256
QY 541 AGAAGATCTAGAACAAAGACAAAGTCAAGGTCAATTTCTCTCACTCACATGATGGTGGTGGT 600
Db 2257 AGAAGATCTAGAACAAAGACAAAGTCAAGGTCAATTTCTCTCACTCACATGATGGTGGTGGT 2316
QY 601 TGATGAATCTAGTGGAGATCACGCAACTGCTTTTGAAGAAACAACTTAAGGTATTTGGG 660
Db 2317 TGATGAATCTAGTGGAGATCACGCAACTGCTTTTGAAGAAACAACTTAAGGTATTTGGG 2376
QY 661 AGATCGATGGGCAACATCTGTAGATGGACAGAAAGACCGGTGGGTTCTTTTACAAGA --- 717
Db 2377 AGATCGATGGGCAACATCTGTAGATGGACAGAAAGACCGGTGGGTTCTTTTACAAGACAC 2436
QY 718 ----- 717
Db 2437 TCATAGATTACTGCAACAGTTCCCTCCCTGGACCTGGAAAGTTTCTTGCTGGCTTACAGA 2496
QY 718 ----- 717
Db 2497 AGCTGAAACAACCTGCCAATGTCTACAGGATGCTACCCGTAAGGAAAGGCTCCTAGAAGA 2556
QY 718 ----- 717
Db 2557 CTCCAAGGAGTAAAGAGCTGTATGAAACAATGGCAAGACCTCCAAGGTGAAATTTGAAGC 2616

QY 718 ----- 717
Db 2617 TCACACAGATGTTTATCACAACCTGGATGAAACAGCCAAAAATCCTGAGATCCCTGGA 2676
QY 718 ----- 717
Db 2677 AGGTTCCGATGATGCAGTCTCTTACAAAGACGTTTGGATAACATGAACCTTCAAGTGGAG 2736
QY 718 -----CAGTTCTGACCAGTG 732
Db 2737 TGAACCTCGGAAAAAGTCTCTCAACATTAGGTCCCATTTGGAAGCCAGTTCTGACCAGTG 2796
QY 733 GAACCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGTGGCTACAGCTGAAAGATGATGA 792
Db 2797 GAACCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGTGGCTACAGCTGAAAGATGATGA 2856
QY 793 ATTAAGCCGGCAGGACCTATTGGAGGCGACTTTCCAGCAGTTTCCAGAGCAGAAACGATGT 852
Db 2857 ATTAAGCCGGCAGGACCTATTGGAGGCGACTTTCCAGCAGTTTCCAGAGCAGAAACGATGT 2916
QY 853 ACATAGGGCCTTCAAGAGGGGAATTTGAAAACTTAAAGAACCTGTAAATCATGAGTACTCTTGA 912
Db 2917 ACATAGGGCCTTCAAGAGGGGAATTTGAAAACTTAAAGAACCTGTAAATCATGAGTACTCTTGA 2976
QY 913 GACTGTACGAATATTTCTGACAGAGCAGCCTTTTGGAAAGGACTAGAGAAACTCTACCAGGA 972
Db 2977 GACTGTACGAATATTTCTGACAGAGCAGCCTTTTGGAAAGGACTAGAGAAACTCTACCAGGA 3036
QY 973 GCCCAGAGAGCTGCCCTCTGAGGAGAGAGAGCCCAAGAAATGTCACCTCGGCTTCTACGAAAGCA 1032
Db 3037 GCCCAGAGAGCTGCCCTCTGAGGAGAGAGAGCCCAAGAAATGTCACCTCGGCTTCTACGAAAGCA 3096
QY 1033 GGCTGAGGAGTCAATACTACTGAGTGGGAAAAATTTGAACCTGCACCTCGGCTGAGTGGCAGAG 1092
Db 3097 GGCTGAGGAGTCAATACTACTGAGTGGGAAAAATTTGAACCTGCACCTCGGCTGAGTGGCAGAG 3156
QY 1093 AAAAAATAGATGAGACCTCTTGAAGAGACTCCAGGAACTTCAAGAGGCCACGGATGAGCTGGA 1152
Db 3157 AAAAAATAGATGAGACCTCTTGAAGAGACTCCAGGAACTTCAAGAGGCCACGGATGAGCTGGA 3216
QY 1153 CCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATCTTGGCAGCCCGTGGCGGATCTCCT 1212
Db 3217 CCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATCTTGGCAGCCCGTGGCGGATCTCCT 3276
QY 1213 CATTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGACTTCAAGGAGAAATTTGCGCC 1272
Db 3277 CATTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGACTTCAAGGAGAAATTTGCGCC 3336
QY 1273 TCTGAAAGAGAACGTGAGCCACGTCAATGACTTGGAAAGACCTGAACACCCAGATGGAAGCTTCT 1332
Db 3337 TCTGAAAGAGAACGTGAGCCACGTCAATGACTTGGAAAGACCTGAACACCCAGATGGAAGCTTCT 3396
QY 1333 TCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCTGCAATGACTTGGAAAGACCTGAACACCCAGATGGAAGCTTCT 1392
Db 3397 TCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCTGCAATGACTTGGAAAGACCTGAACACCCAGATGGAAGCTTCT 3456
QY 1393 GCAGGTGGCCGTGAGGACCGAGTCAAGCAGCTGCAATGACTTGGAAAGACCTGAACACCCAGATGGAAGCTTCT 1452
Db 3457 GCAGGTGGCCGTGAGGACCGAGTCAAGCAGCTGCAATGACTTGGAAAGACCTGAACACCCAGATGGAAGCTTCT 3516
QY 1453 AGCATCTCAGCACTTTCTTCCACGTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCC 1512
Db 3517 AGCATCTCAGCACTTTCTTCCACGTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCC 3576
QY 1513 AAACAAAGTGCCCTACTATATCAACACAGAGACTCAAAACAACCTTCTGCGGACCATCCCAA 1572
Db 3577 AAACAAAGTGCCCTACTATATCAACACAGAGACTCAAAACAACCTTCTGCGGACCATCCCAA 3636
QY 1573 AATGACAGAGCTTACCAGTCTTTAGCTGACCTGAATAATGTCAGATTTCTCAGCTTATAG 1632
Db 3637 AATGACAGAGCTTACCAGTCTTTAGCTGACCTGAATAATGTCAGATTTCTCAGCTTATAG 3696
QY 1633 GACTGCCATGAAACTCCGAAAGACTGCAGAAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTC 1692

||||| 3697 GACTGCCATGAAACCTCCGAGACTGCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTC 3756
1693 AGCTGCATGTGATGCCCTTGGACAGCACAACTCAAGCAAAATGACAGCCCATGGATAT 1752
3757 AGCTGCATGTGATGCCCTTGGACAGCACAACTCAAGCAAAATGACAGCCCATGGATAT 3816
1753 CCTGCAGATTATTAATTGTTTGACCACTATTATGACCCGCTGGAGCAAGAGCACAA 1812
3817 CCTGCAGATTATTAATTGTTTGACCACTATTATGACCCGCTGGAGCAAGAGCACAA 3876
1813 TTTGGTCAACGTCCTCTCTGCGTGGATATGTCTGAACTGGCTGCTGAATGTTTATGA 1872
3877 TTTGGTCAACGTCCTCTCTGCGTGGATATGTCTGAACTGGCTGCTGAATGTTTATGA 3936
1873 TACGGGACGAACAGGGAGGATCCGTGCTGCTTTTAAACTGGCATCATTTCCCTGTG 1932
3937 TACGGGACGAACAGGGAGGATCCGTGCTGCTTTTAAACTGGCATCATTTCCCTGTG 3996
1933 TAAAGCACATTTGGAAGACAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGG 1992
3997 TAAAGCACATTTGGAAGACAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGG 4056
1993 ATTTTGTGACCGCAGGCTGGGCTCCTTCTGCTGATTTCTATCCAA 2041
4057 ATTTTGTGACCGCAGGCTGGGCTCCTTCTGCTGATTTCTATCCAA 4105
RESULT 7
AAD37263
ID AAD37263 standard; DNA; 4848 BP.
XX
AC AAD37263;
XX
DT 21-AUG-2002 (first entry)
XX
DE Adeno-associated virus (AAV) vector plasmid, AAV-CMV-delta3849.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Cytomegalovirus.
OS Chimeric - Unidentified.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
XX
PI Xiao X;
XX
PI WPI; 2002-049342/06.
XX
PT New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene
XX
PS Example 1; Page 68-70; 71pp; English.
XX
CC The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control

CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is AAV
CC vector plasmid construct containing human dystrophin minigenes, a
CC cytomagalovirus (CMV) promoter and a small polyA signal sequence.
XX
SQ Sequence 4848 BP; 1405 A; 1144 C; 1180 G; 1119 T; 0 other;
Query Match 82.5%; Score 1683; DB 24; Length 4848;
Best Local Similarity 85.4%; Pred. No. 0;
Matches 2041; Conservative 0; Mismatches 0; Indels 348; Gaps 1;
QY 1 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAAGTTCATTGGCAGTTTCATTGATGGAGAG 60
Db 1740 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAAGTTCATTGGCAGTTTCATTGATGGAGAG 1799
QY 61 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGATATTATCGTGGCTTCTTTC 120
Db 1800 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGATATTATCGTGGCTTCTTTC 1859
QY 121 TGCTGAGGACACATTTGCAAGCACAAAGGAGAGATTCTTAATGATGTGGAAGTGGTGAAGA 180
Db 1860 TGCTGAGGACACATTTGCAAGCACAAAGGAGAGATTCTTAATGATGTGGAAGTGGTGAAGA 1919
QY 181 CCAGTTTCATCTACTCATGAGGGGTACATGATGATTTGACAGCCCATCAGGGCCGGTTGG 240
Db 1920 CCAGTTTCATCTACTCATGAGGGGTACATGATGATTTGACAGCCCATCAGGGCCGGTTGG 1979
QY 241 TAATATTCTACAATTGSGAAGTAAGCTGATTGGAACAGGAAAAATATCAGAAGATGAAGA 300
Db 1980 TAATATTCTACAATTGSGAAGTAAGCTGATTGGAACAGGAAAAATATCAGAAGATGAAGA 2039
QY 301 AACTGAAGTACAAGAGCAGATGAATCTCTAAATTCGAAGTGGGAATGCCTCAGGGTAGC 360
Db 2040 AACTGAAGTACAAGAGCAGATGAATCTCTAAATTCGAAGTGGGAATGCCTCAGGGTAGC 2099
QY 361 TAGCATGGAACAAACAAAGCAATTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT 420
Db 2100 TAGCATGGAACAAACAAAGCAATTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT 2159
QY 421 GAAAGAGTTGAATGACTGGCTTAACAAAAACAGAAAGAAACAAAGGAAAAATGGAGGAAGA 480
Db 2160 GAAAGAGTTGAATGACTGGCTTAACAAAAACAGAAAGAAACAAAGGAAAAATGGAGGAAGA 2219
QY 481 GCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCCAAGTACAACAACTAAGGTGCTTCA 540
Db 2220 GCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCCAAGTACAACAACTAAGGTGCTTCA 2279
QY 541 AGAAGATCTAGAACAAGAACAAAGTCAGGGTCAATTCTCTCACTCACATGGTGGTGGTAGT 600
Db 2280 AGAAGATCTAGAACAAGAACAAAGTCAGGGTCAATTCTCTCACTCACATGGTGGTGGTAGT 2339
QY 601 TGATGAATCTAGTGGAGATCAGGCAACTGCTTGGGAAGAACAACTTAAGGTATTGGG 660
Db 2340 TGATGAATCTAGTGGAGATCAGGCAACTGCTTGGGAAGAACAACTTAAGGTATTGGG 2399
QY 661 AGATCGATGGGCAAAACATCTGTAGATGGACAGACCCGCTGGTTCCTTTTACAAGA--- 717
Db 2400 AGATCGATGGGCAAAACATCTGTAGATGGACAGACCCGCTGGTTCCTTTTACAAGACAC 2459
QY 718 ----- 717
Db 2460 TCATAGATTACTGCAACAGTTCCCTTGGACCTGGAAAGATTCTTGCCTGGCTTACAGA 2519
QY 718 ----- 717
Db 2520 AGCTGAAACAACACTGCCAATGTCTACAGGATGCTACCCGTAAGGAAGGCTCCTAGAAGA 2579
QY 718 ----- 717
Db 2580 CTCCAAGGGAGTAAAGAGCTGATGAAACAATGGCAAGACCTCCAAGGTGAAATTGAAGC 2639

QY 718 ----- 717
Db 2640 TCACACAGATGTTTATCACAACTGGATGAAACAGCCAAAAAATCCTGAGATCCCTGGA 2699
QY 718 ----- 717
Db 2700 AGGTTCCGATGATGCAGTCCTGTTACAAAGACGTTTGGATAACATGAACCTTCAAGTGGAG 2759
QY 718 -----CAGTTCTGACCAGTG 732
Db 2760 TGAACCTCGGAAAAAGTCTCTCAACATTAGGTCCCATTTGGAAGCCAGTTCTGACCAGTG 2819
QY 733 GAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGTGGCTACAGCTGAAAGATGATGA 792
Db 2820 GAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGTGGCTACAGCTGAAAGATGATGA 2879
QY 793 ATTAAGCCGGCAGGCACCTATTGGAGGCGACTTTCAGCAGCTTCAGAAAGCAGAACCGATGT 852
Db 2880 ATTAAGCCGGCAGGCACCTATTGGAGGCGACTTTCAGCAGCTTCAGAAAGCAGAACCGATGT 2939
QY 853 ACATAGGGCCCTTCAAGAGGGAATTGAAAACTAAAGAACCTGTAAATCATGAGTACTCTTGA 912
Db 2940 ACATAGGGCCCTTCAAGAGGGAATTGAAAACTAAAGAACCTGTAAATCATGAGTACTCTTGA 2999
QY 913 GACTGTACGAATATTCTGCAGAGCAGCCCTTTGGAAGGACTAGAGAAACTCTACCCAGGA 972
Db 3000 GACTGTACGAATATTCTGCAGAGCAGCCCTTTGGAAGGACTAGAGAAACTCTACCCAGGA 3059
QY 973 GCCCAGAGAGCTGCCTCCTGAGGAGAGAGCCCCAGAAATGTCACTCGGCTTCTACGAAAAAGCA 1032
Db 3060 GCCCAGAGAGCTGCCTCCTGAGGAGAGAGAGCCCCAGAAATGTCACTCGGCTTCTACGAAAAAGCA 3119
QY 1033 GGCTGAGGAGGTCAATACTAGTGGGAAAAAATTGAACCTGCACCTCCGCTGACTGGCAGAG 1092
Db 3120 GGCTGAGGAGGTCAATACTAGTGGGAAAAAATTGAACCTGCACCTCCGCTGACTGGCAGAG 3179
QY 1093 AAAAATAGATGAGACCCCTTGAAGACTCCAGGAACCTTCAAGAGGCCACGGATGAGCTGGA 1152
Db 3180 AAAAATAGATGAGACCCCTTGAAGACTCCAGGAACCTTCAAGAGGCCACGGATGAGCTGGA 3239
QY 1153 CCTCAAGCTGCGCCAAAGTGAAGTGTATCAAGGGATCCTGGCAGCCCGTGGCGGATCTCCT 1212
Db 3240 CCTCAAGCTGCGCCAAAGTGAAGTGTATCAAGGGATCCTGGCAGCCCGTGGCGGATCTCCT 3299
QY 1213 CATTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAAAATTGCGCC 1272
Db 3300 CATTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAAAATTGCGCC 3359
QY 1273 TCTGAAAGAGAACGTGAGCCACGTCATGACCTTGCTCGCCAGCTTACCACCTTTGGGCAT 1332
Db 3360 TCTGAAAGAGAACGTGAGCCACGTCATGACCTTGCTCGCCAGCTTACCACCTTTGGGCAT 3419
QY 1333 TCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCTGAACACCCAGATGGAAGCTTCT 1392
Db 3420 TCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCTGAACACCCAGATGGAAGCTTCT 3479
QY 1393 GCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGAAGCCCAACAGGGACTTTGGTCC 1452
Db 3480 GCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGAAGCCCAACAGGGACTTTGGTCC 3539
QY 1453 AGCATCTCAGCACTTTCTTCCACGTCGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCC 1512
Db 3540 AGCATCTCAGCACTTTCTTCCACGTCGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCC 3599
QY 1513 AAACAAAGTGCCTACTATATCAACCCAGAGACTCAAAACAACTTGCTGGGACCATCCCAA 1572
Db 3600 AAACAAAGTGCCTACTATATCAACCCAGAGACTCAAAACAACTTGCTGGGACCATCCCAA 3659
QY 1573 AATGACAGAGCTCTACCAAGTCTTTAGCTGACCTGAATAATGTGAGATTCTCAGCTTATAG 1632
Db 3660 AATGACAGAGCTCTACCAAGTCTTTAGCTGACCTGAATAATGTGAGATTCTCAGCTTATAG 3719
QY 1633 GACTGCCATGAAACTCCGAAGACTGCAGAAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTC 1692

Db 3720 GACTGCCATGAAACTCCGAAGACTGCAGAAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTC 3779
QY 1693 AGCTGCATGTGATGCCTTGACCAGCACAAACCTCAAGCAAAATGACCAGCCCATGGATAT 1752
Db 3780 AGCTGCATGTGATGCCTTGACCAGCACAAACCTCAAGCAAAATGACCAGCCCATGGATAT 3839
QY 1753 CCTGCAGATTATTAAATTGTTGACCACACTATTATGACCCGCTGGAGCAAGAGCACAA 1812
Db 3840 CCTGCAGATTATTAAATTGTTGACCACACTATTATGACCCGCTGGAGCAAGAGCACAA 3899
QY 1813 TTTGGTCAACGTCCTCTCTGCGTGGATATGTGTCTGAACTGGCTGCTGAATGTTTATGA 1872
Db 3900 TTTGGTCAACGTCCTCTCTGCGTGGATATGTGTCTGAACTGGCTGCTGAATGTTTATGA 3959
QY 1873 TACGGGACGAACAGGGAGGATCCGTGTCTCTCTTTTAAAACTGGCATCATTTCCCTGTG 1932
Db 3960 TACGGGACGAACAGGGAGGATCCGTGTCTCTCTTTTAAAACTGGCATCATTTCCCTGTG 4019
QY 1933 TAAAGCACATTTGGAAGACAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGG 1992
Db 4020 TAAAGCACATTTGGAAGACAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGG 4079
QY 1993 ATTTTGTGACCAGCGCAGGCTGGGCTCTCTTCTGTCATGATTCTATCCAA 2041
Db 4080 ATTTTGTGACCAGCGCAGGCTGGGCTCTCTTCTGTCATGATTCTATCCAA 4128

RESULT 8

AAD37260

ID AAD37260 standard; DNA; 4414 BP.

XX AAD37260;
AC AAD37260;

XX 21-AUG-2002 (first entry)

Adeno-associated virus vector plasmid, AAV-MCK-3447.

Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
Becker muscular dystrophy; ds.

Chimeric - Homo sapiens.
Chimeric - Unidentified.

WO200183695-A2.

08-NOV-2001.

27-APR-2001; 2001WO-US13677.

28-APR-2000; 2000US-200777P.

(XIAO/) XIAO X.

Xiao X;

WPI; 2002-049342/06.

New dystrophin minigene for treating Duchenne or Becker muscular
dystrophy comprises an N-terminal domain or modified N-terminal domain,
rod repeats, H1 and H4 domains and a cysteine rich domain of a
dystrophin gene -

Example 1; Page 65-66; 71pp; English.

The present invention relates to an isolated nucleotide sequence encoding
a dystrophin minigene. The minigene comprises N-terminal or modified
N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
domains and cysteine-rich domains of dystrophin or utrophin genes. The
invention also relates to a recombinant adeno-associated virus (AAV)
comprising dystrophin minigene operably linked to an expression control
element. The dystrophin minigene in operable linkage with an expression

control element, in a recombinant adeno-associated virus or retrovirus is useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is AAV vector plasmid construct containing human dystrophin minigenes, a muscle creatine kinase (MCK) promoter and a small polvA signal sequence.

Sequence 4414 BP: 1255 A; 1075 C; 1086 G; 998 T; 0 other;

Query Match 77.5%; Score 1581; DB 24; Length 4414;
Best Local Similarity 87.6%; Pred. No. 0;
Matches 1788; Conservative 0; Mismatches 190; Indels 63; Gaps 3;

1	TCCTTCACAGCATTTGGAGCTCCTGAAGACAAGTCATTGGCAGTTTCATTGATGGAGAG	60
1717	TCCTTCACAGCATTTGGAGCTCCTGAAGACAAGTCATTGGCAGTTTCATTGATGGAGAG	1776
61	TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAGAAGPATATCGTGGCTTCTTTC	120
1777	TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAGAAGPATATCGTGGCTTCTTTC	1836
121	TGCTGAGGACACATTGCAAGCACAAGGAGAGATTTCATATGATGCGAAAGTGGTGAAGA	180
1837	TGCTGAGGACACATTGCAAGCACAAGGAGAGATTTCATATGATGCGAAAGTGGTGAAGA	1896
181	CCAGTTTCATACTCATGAGGGGTACATGATGGATTTGACAGCCCATCAGGGCCGGTTGG	240
1897	CCAGTTTCATACTCATGAGGGGTACATGATGGATTTGACAGCCCATCAGGGCCGGTTGG	1956
241	TAATATTCTACAATTGGGAAGTAAGCTGATTGGAACAGGAAAAATTTATCAGAAGATGAAGA	300
1957	TAATATTCTACAATTGGGAAGTAAGCTGATTGGAACAGGAAAAATTTATCAGAAGATGAAGA	2016
301	AACCTGAAGTACAAGACGAGATGAATCTCCTAAATTCAGAGATGGGAATGCCTCAGGGTAGC	360
2017	AACCTGAAGTACAAGACGAGATGAATCTCCTAAATTCAGAGATGGGAATGCCTCAGGGTAGC	2076
361	TAGCATGGAAAAACAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAAACT	420
2077	TAGCATGGAAAAACAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAAACT	2136
421	GAAAGATTGAATGACTGGCTAACAAAAACAGAAAGAAACAAGGAAATTCGAGGAAGA	480
2137	GAAAGATTGAATGACTGGCTAACAAAAACAGAAAGAAACAAGGAAATTCGAGGAAGA	2196
481	GCCTCTTGGACCTGATCTTGAAGACCTAAACGCCCAAGTACAACAACATAAGTGCTTCA	540
2197	GCCTCTTGGACCTGATCTTGAAGACCTAAACGCCCAAGTACAACAACATAAGTGCTTCA	2256
541	AGAAGATCTAGAACAAGAACAAGTCAGGGTCAATTTCTCACTCACATGGTGGTGGTAGT	600
2257	AGAAGATCTAGAACAAGAACAAGTCAGGGTCAATTTCTCACTCACATGGTGGTGGTAGT	2316
601	TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTTGGAAAGAACAACCTTAAGGTAATGGG	660
2317	TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTTGGAAAGAACAACCTTAAGGTAATGGG	2376
661	AGATCGATGGGCAAAACATCTGTAGATGGACAGAGACCCGCTGGGTTCTTTTACAAGACAG	720
2377	AGATCGATGGGCAAAACATCTGTAGATGGACAGAGACCCGCTGGGTTCTTTTACAAGACAT	2436
721	TTCTGACCAGTGGAAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGTGGCTACAGCT	780
2437	CCTTCTCAAAATGGCAACGCTCTTACTGAAGAACAACAGTGCCTTTTGTAGTCATGGCTTTCAGA	2496
781	GAAAGATGATGAATTAAGCCGGCAGGCACCTATTGGAGGCGACTTTCCAGCAGCTTCAGAA	840
2497	AAAAGAAGATGCAGTGAACAAGATTCCACACAACCTGGCTTTTAAAGATCAAAATGAAATGTT	2556
841	GCAGAACGATGTACATAGGCGCTTCAAGAGGGAATTGAAAACCTAAAGAACCTGTAATCAT	900
2557	ATCAAGTCTTCAAAAACCTGGCCGCTTTTAAAGCGGATCTAGAAAAGAAAAAGCAATCCAT	2616
901	GAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTTGAAGGACTAGAGAA	960

Db 3634 AAGTTCAACAGGATTTGTGACACGCGAGGCTGGGCTCCTTCTGTCATGATTTCTATCCA 3693

QY 2041 A 2041

Db 3694 A 3694

RESULT 9

AAD37242

ID AAD37242 standard; DNA; 3446 BP.

XX AC AAD37242;

XX DT 21-AUG-2002 (first entry)

DE Human dystrophin minigene delta3447.

XX KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;

KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;

XX KW Becker muscular dystrophy; ds.

OS Homo sapiens.

XX PN WO200183695-A2.

XX PD 08-NOV-2001.

XX PF 27-APR-2001; 2001WO-US13677.

XX PR 28-APR-2000; 2000US-200777P.

XX PA (XIAO/) XIAO X.

XX XIAO X;

XX WPI; 2002-049342/06.

PT New dystrophin minigene for treating Duchenne or Becker muscular

PT dystrophy comprises an N-terminal domain or modified N-terminal domain,

PT rod repeats, H1 and H4 domains and a cysteine rich domain of a

PT dystrophin gene -

XX Example 1; Page 53-54; 71pp; English.

XX The present invention relates to an isolated nucleotide sequence encoding

CC a dystrophin minigene. The minigene comprises N-terminal or modified

CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4

CC domains and cysteine-rich domains of dystrophin or utrophin genes. The

CC invention also relates to a recombinant adeno-associated virus (AAV)

CC comprising dystrophin minigene operably linked to an expression control

CC element. The dystrophin minigene in operable linkage with an expression

CC control element, in a recombinant adeno-associated virus or retrovirus is

CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular

CC dystrophy (BMD) in a mammalian subject. The present sequence is human

CC dystrophin minigene delta3447 containing nucleotides 1-1992 (N-terminus,

CC hinge H1 and rods R1, R2 and R3), 8749-10227 (rod R24, hinge H4 and

CC CR domain) and 11047-11058 (dystrophin last 3 amino acids).

XX Sequence 3446 BP; 1074 A; 766 C; 787 G; 819 T; 0 other;

SQ

Query Match 76.9%; Score 1569; DB 24; Length 3446;

Best Local Similarity 87.6%; Pred. No. 0;

Matches 1787; Conservative 0; Mismatches 190; Indels 64; Gaps 4;

QY 1 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAAGTCATTTGGCAGTTTCATTGAGAGAG 60

Db 960 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAAGTCATTTGGCAGTTTCATTGAGAGAG 1019

QY 61 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAGAAGTATTATCGTGGCTTCTTTC 120

Db 1020 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAGAAGTATTATCGTGGCTTCTTTC 1079

QY 121 TGCTGAGGACACATTTGCAAGCACAAAGGAGAGATTTCTTAATGATGTGGAAGTGGTGAAGA 180

Db 1080 TGCTGAGGACACATTTGCAAGCACAGGAGAGATTTCTTAATGATGTGGAAGTGGTGAAGA 1139

QY 181 CCAGTTTCATCTACTGAGGGGTACATGATGATGATGATGATGATGATGATGATGATGATGATG 240

Db 1140 CCAGTTTCATCTACTGAGGGGTACATGATGATGATGATGATGATGATGATGATGATGATG 1199

QY 241 TAATATTTACAATTTGGGAAGTAAAGCTGATGATGATGATGATGATGATGATGATGATGATG 300

Db 1200 TAATATTTACAATTTGGGAAGTAAAGCTGATGATGATGATGATGATGATGATGATGATGATG 1259

QY 301 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTTCAAGATGGGAATGCTCAGGGTAGC 360

Db 1260 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTTCAAGATGGGAATGCTCAGGGTAGC 1319

QY 361 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAAACT 420

Db 1320 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATC-GAAACT 1378

QY 421 GAAAGAGTTGAATGACTGGCTAAACAAAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 480

Db 1379 GAAAGAGTTGAATGACTGGCTAAACAAAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 1438

QY 481 GCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCAAGTACAACAAACATAAGGTGCTTCA 540

Db 1439 GCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCAAGTACAACAAACATAAGGTGCTTCA 1498

QY 541 AGAAGATCTAGAACAAAGCAAGTCAAGGTCAATTTCTCTCACTCACATGGTGGTAGT 600

Db 1499 AGAAGATCTAGAACAAAGCAAGTCAAGGTCAATTTCTCTCACTCACATGGTGGTAGT 1558

QY 601 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTTGAAGAACAACACTTAAGGTATGGG 660

Db 1559 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTTGAAGAACAACACTTAAGGTATGGG 1618

QY 661 AGATCGATGGGCAAAACATCTGTAGATGGACAGACCCGCTGGGTTCTTTTACAAGACAG 720

Db 1619 AGATCGATGGGCAAAACATCTGTAGATGGACAGACCCGCTGGGTTCTTTTACAAGACAT 1678

QY 721 TTCTGACCAAGTGGAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGGCTACAGCT 780

Db 1679 CCTTCTCAAATGGCAACGCTTTACTGAAGAACAAGTGCCTTTTGTAGTGCATGGCTTTCAGA 1738

QY 781 GAAAGATGATGAATTAAGCCGGCAGGACACCTTATTTGGAGGGGACTTTTCAGACAGTTTCAGAA 840

Db 1739 AAAAGAAGATGCAGTGAACAAGATTTACACAACTGGCTTTAAAGATCAAAATGAAATGTT 1798

QY 841 GCAGAACGATGTACATAGGCGCTTCAAGAGGGAATTTGAAACCTTAAAGAACCTGTATCAT 900

Db 1799 ATCAAGTCTTCAAAACACTGGCGCTTTTAAAGCGGATCTAGAAAAGAAAAGCAATCCAT 1858

QY 901 GAGTACTCTTGAGACTCTACGAATATTTCTGACAGAGCAGCTTTTGAAGGAGACTAGAGAA 960

Db 1859 GGGCAAACTGTA-----TTCACTCAAACAAGATCTTCTTTCAACACTGAAG 1904

QY 961 ACTCTACAGGAGCCAGAGAGCTGCCCTCCTGAGGAGAGAGCCCAAGATGTCACTCGGCT 1020

Db 1905 AATAAGTCAAGTGACCCAGAGACGGAAGCATGGCTGGA----- 1942

QY 1021 TCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGGAAGAAATTTGAACCTGCACCTCCGC 1080

Db 1943 -----TAACCTTGGCCCGGTGTTGGGATAATTAGTCCAAAAAACTTGAA----- 1985

QY 1081 TGACTGGCAGAGAAAAATAGATGAGACCTTGAAAGACTCCAGGAACCTTCAAGAGGCCAC 1140

Db 1986 -----AAGGTACAGCACAGACCTTGAAAGACTCCAGGAACCTTCAAGAGGCCAC 2035

QY 1141 GGATGAGCTGGACCTCAAGCTGGCCCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGT 1200

Db 2036 GGATGAGCTGGACCTCAAGCTGGCCCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGT 2095

QY 1201 GGGCGATCTCCTCATTTGACTCTTCCAAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGG 1260

Db 2096 GGGGATCTCCTCATTTGACTCTCTCCAAAGATCACCTCGAGAAAGTCAAGCCACTTCGAGG 2155
QY 1261 AGAAATTCGGCCTCTGAAAGAGAACGTCAGCCACGTCATGACCTTGTCTGCCAGCTTAC 1320
Db 2156 AGAAATTCGGCCTCTGAAAGAGAACGTCAGCCACGTCATGACCTTGTCTGCCAGCTTAC 2215
QY 1321 CACTTTGGGCACTTCAGCTCTCACCGTATACCTCAGCACTCTGGAAGACCTGAACACCAG 1380
Db 2216 CACTTTGGGCACTTCAGCTCTCACCGTATACCTCAGCACTCTGGAAGACCTGAACACCAG 2275
QY 1381 ATGGAAGCTTCTGAGGTTGGCGTTCGAGGACCGAGTCCAGGCTGCATGAAGCCACAG 1440
Db 2276 ATGGAAGCTTCTGAGGTTGGCGTTCGAGGACCGAGTCCAGGCTGCATGAAGCCACAG 2335
QY 1441 GGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGTCGTCCAGGCTCCCTGGGAGAG 1500
Db 2336 GGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGTCGTCCAGGCTCCCTGGGAGAG 2395
QY 1501 AGCCATCTCGCCAAACAAAGTGCCTTACTATATCAACCACGAGACTCAACAACTTGCTG 1560
Db 2396 AGCCATCTCGCCAAACAAAGTGCCTTACTATATCAACCACGAGACTCAACAACTTGCTG 2455
QY 1561 GGACCATCCCAAAATGACAGAGCTCTACCACTTTTAGCTGACTGAATAATGTCAGATT 1620
Db 2456 GGACCATCCCAAAATGACAGAGCTCTACCACTTTTAGCTGACTGAATAATGTCAGATT 2515
QY 1621 CTCAGCTTATAGGACTGCCATGAAACTCCGAAAGACTGCAGAAAGCCCTTGTCTGGATCT 1680
Db 2516 CTCAGCTTATAGGACTGCCATGAAACTCCGAAAGACTGCAGAAAGCCCTTGTCTGGATCT 2575
QY 1681 CTTGAGCCTGTCAGCTGCATGTGATGCTTGGACCAAGCAACCTCAAGCAAAATGACCA 1740
Db 2576 CTTGAGCCTGTCAGCTGCATGTGATGCTTGGACCAAGCAACCTCAAGCAAAATGACCA 2635
QY 1741 GCCCATGGATATCCTGCAGATTATTAATTGTTTGACCACTATTTATGACCGCCTGGAGCA 1800
Db 2636 GCCCATGGATATCCTGCAGATTATTAATTGTTTGACCACTATTTATGACCGCCTGGAGCA 2695
QY 1801 AGAGCACAAACAATTTGGTCAACGTCCTCTCTGCTGGATATGTGTCTGAACCTGGCTGCT 1860
Db 2696 AGAGCACAAACAATTTGGTCAACGTCCTCTCTGCTGGATATGTGTCTGAACCTGGCTGCT 2755
QY 1861 GAATGTTTATGATACGGGACGAAACAGGAGGATCCGTCCTCTCTTTTAAACCTGGCAT 1920
Db 2756 GAATGTTTATGATACGGGACGAAACAGGAGGATCCGTCCTCTCTTTTAAACCTGGCAT 2815
QY 1921 CATTTCCTGTGTAAGACACATTTGGAAGACAAAGTACAGATACCTTTTCAAGCAAGTGGC 1980
Db 2816 CATTTCCTGTGTAAGACACATTTGGAAGACAAAGTACAGATACCTTTTCAAGCAAGTGGC 2875
QY 1981 AAGTTCAACAGGATTTTGTGACCAAGCAGGCTGGGCTCCTCTCTGCTGATGATTTCTATCCA 2040
Db 2876 AAGTTCAACAGGATTTTGTGACCAAGCAGGCTGGGCTCCTCTCTGCTGATGATTTCTATCCA 2935
QY 2041 A 2041
Db 2936 A 2936
RESULT 10
AAD37234
ID AAD37234 standard; DNA; 3999 BP.
XX
AC AAD37234;
XX
DT 21-AUG-2002 (first entry)
XX
DE Human dystrophin minigene delta3990.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX

Homo sapiens.
WO200183695-A2.
08-NOV-2001.
27-APR-2001; 2001WO-US13677.
28-APR-2000; 2000US-200777P.
(XIAO/) XIAO X.
xiao X;
WPI; 2002-049342/06.
New dystrophin minigene for treating Duchenne or Becker muscular dystrophy comprises an N-terminal domain or modified N-terminal domain, rod repeats, H1 and H4 domains and a cysteine rich domain of a dystrophin gene -
Example 1; Page 46-47; 71pp; English.
The present invention relates to an isolated nucleotide sequence encoding a dystrophin minigene. The minigene comprises N-terminal or modified N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control element. The dystrophin minigene in operable linkage with an expression control element, in a recombinant adeno-associated virus or retrovirus is useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is human dystrophin minigene delta3990 containing nucleotides 1-1668 (N-terminus, hinge H1 and rods R1 and R2), 8059-10227 (rods R22, R23 and R24, hinge H4 and CR domain) and 11047-11058 (dystrophin last 3 amino acids).
Sequence 3999 BP; 1223 A; 907 C; 933 G; 936 T; 0 other;

Query Match 75.6%; Score 1542; DB 24; Length 3999;
Best Local Similarity 80.7%; Pred. No. 0;
Matches 2041; Conservative 0; Mismatches 0; Indels 489; Gaps 1;
QY 1 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAAGTCAATTTGGCAGTTCATTGATGGAGAG 60
Db 960 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAAGTCAATTTGGCAGTTCATTGATGGAGAG 1019
QY 61 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAGTATATCGTGGCTCTTTC 120
Db 1020 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAGTATATCGTGGCTCTTTC 1079
QY 121 TGCTGAGGACACATTTGCAAGCAAGGAGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 180
Db 1080 TGCTGAGGACACATTTGCAAGCAAGGAGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 1139
QY 181 CCAGTTTCATACACTCATGAGGGGTACATGATGGATTTGACAGCCCATCAGGCGCGGTTGG 240
Db 1140 CCAGTTTCATACACTCATGAGGGGTACATGATGGATTTGACAGCCCATCAGGCGCGGTTGG 1199
QY 241 TAATATTCTACAAATTTGGGAAGTAAGCTGATTTGGAACAGGAAATTTATCAGAAGATGAAGA 300
Db 1200 TAATATTCTACAAATTTGGGAAGTAAGCTGATTTGGAACAGGAAATTTATCAGAAGATGAAGA 1259
QY 301 AACTGAAGTACAAGAGCAGATGAATCTCTCTAAATTCAGATGGGAATGCCTCAGGCTAGC 360
Db 1260 AACTGAAGTACAAGAGCAGATGAATCTCTCTAAATTCAGATGGGAATGCCTCAGGCTAGC 1319
QY 361 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAAACT 420
Db 1320 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAAACT 1379
QY 421 GAAAGAGTTGAATGACTGGCTAACAAAAACAGAGAAAGAAACAAAGGAAATGGAGGAAGA 480

Db	1380	GAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGAAACAAGGAAAAATGGAGGAAGA	1433
QY	481	GCCTCTTGGACCTGATCTTGAAGACCTAAAAACGCCAAGTACAAACAACATAAGGTGCTTCA	540
Db	1440	GCCTCTTGGACCTGATCTTGAAGACCTAAAAACGCCAAGTACAAACAACATAAGGTGCTTCA	1499
QY	541	AGAAGATCTAGAACAAAGAACAAAGTCAGGGTCAATTCTCTCACTCACATGGTGGTAGT	600
Db	1500	AGAAGATCTAGAACAAAGAACAAAGTCAGGGTCAATTCTCTCACTCACATGGTGGTAGT	1559
QY	601	TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTTGGAAAGAACAACTTAAGGTATTGGG	660
Db	1560	TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTTGGAAAGAACAACTTAAGGTATTGGG	1619
QY	661	AGATCGATGGGCAAAACATCTGTAGATGGACAGAAAGACCGCTGGGTCTTTTACAAGA	717
Db	1620	AGATCGATGGGCAAAACATCTGTAGATGGACAGAAAGACCGCTGGGTCTTTTACAAGACCA	1679
QY	718	-----	717
Db	1680	GCCTGACCTAGCTCCTGGACTGACCACTATTGGAGCCTCTCCTACTCAGACTGTTACTCT	1739
QY	718	-----	717
Db	1740	GGTGACAAACCTGTGGTTACTTAAGGAAACTGCCATCTCCAAACTAGAAATGCCATCTTC	1799
QY	718	-----	717
Db	1800	CTTGATGTTGGAGGTACCTACTCATAGATTACTGCAACAGTTCGCCCTGGACCTGCAAAA	1859
QY	718	-----	717
Db	1860	GTTTCTTGCCCTGGCTTACAGAAAGCTGAAACAACCTGCCAATGTCTCTACAGGATGCTACCCG	1919
QY	718	-----	717
Db	1920	TAAGGAAAGGCTCCTAGAAGACTCCAAGGGAGTAAAGAGAGCTGATGAAACAATGGCAAGA	1979
QY	718	-----	717
Db	1980	CCTCCAAGGTGAAATTGAAGCTCACAGATGTTTATCACAACTGGATGAAAACAGCCA	2039
QY	718	-----	717
Db	2040	AAAAATCCTGAGATCCCTGGAAAGTTCCGATGATGCAGTCTCTGTTACAAAGACGTTTGA	2099
QY	718	-----	717
Db	2100	TAACATGAACCTCAAGTGGAGTGAACCTTCGAAAAAGTCTCTCAACATTAGGTCCCATTT	2159
QY	718	-----CAGTTCTGACCAGTGGAAAGCGTCTGCACCTTTCTCTGCAGGAACCTCTGGTGTG	771
Db	2160	GGAAGCCAGTCTTGACCAGTGGAAAGCGTCTGCACCTTTCTCTGCAGGAACCTCTGGTGTG	2219
QY	772	GCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTGGAGGCGACTTCCAGC	831
Db	2220	GCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTGGAGGCGACTTCCAGC	2279
QY	832	AGTTCAGAAAGCAGAACGATGTACATAGGGCCTTCAAGAGGGAATTGAAAACATAAAGAAC	891
Db	2280	AGTTCAGAAAGCAGAACGATGTACATAGGGCCTTCAAGAGGGAATTGAAAACATAAAGAAC	2339
QY	892	TGTAATCATGAGTACTCTTGAGACTGTACGAATATTCTGACAGAGCAGCCTTTTGAAGG	951
Db	2340	TGTAATCATGAGTACTCTTGAGACTGTACGAATATTCTGACAGAGCAGCCTTTTGAAGG	2399
QY	952	ACTAGAGAAACTCTACCAGGAGCCACAGAGAGCTGCCTCTGTAGGAGAGAGCCCAAGATGT	1011
Db	2400	ACTAGAGAAACTCTACCAGGAGCCACAGAGAGCTGCCTCTGTAGGAGAGAGCCCAAGATGT	2459
QY	1012	CACCTGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTGAACCT	1071
Db	2460	CACCTGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTGAACCT	2519

RESULT 11
AAD37256
ID AAD37256

ID AAD37256 standard; DNA; 4966 BP.

XX AAD37256;
XX
XX
XX DT 21-AUG-2002 (first entry)
XX Adeno-associated virus vector plasmid, AAV-MCK-delta3990.
XX
XX Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
XX Chimeric - Homo sapiens.
OS Chimeric - Unidentified.
OS
XX WO200183695-A2.
PN
XX
XX 08-NOV-2001.
PD
XX
XX 27-APR-2001; 2001WO-US13677.
PF
XX
XX 28-APR-2000; 2000US-200777P.
PR
XX (XIAO/) XIAO X.
PA
XX
XX Xiao X;
PI
XX
XX WPI; 2002-049342/06.
DR
XX
XX New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -
XX
XX Example 1; Page 59-60; 71pp; English.
PS
XX
XX The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is AAV
CC vector plasmid construct containing human dystrophin minigenes, a muscle
CC creatine kinase (MCK) promoter and a small polyA signal sequence.
XX
XX Sequence 4966 BP; 1403 A; 1216 C; 1232 G; 1115 T; 0 other;
SQ
Query Match 75.6%; Score 1542; DB 24; Length 4966;
Best Local Similarity 80.7%; Pred. No. 0;
Matches 2041; Conservative 0; Mismatches 0; Indels 489; Gaps 1;
QY 1 TCCTTCACAGCATTTGGAGCTCCTGAAGACAAAGTCATTTTGGCAGTTCATTGATGGAGAG 60
Db
QY 61 TGAAGTAAACCTGGACCGTTATCAACACAGCTTTAGAGAAGATATATATCGTGGCTTCTTC 120
Db 1777 TGAAGTAAACCTGGACCGTTATCAACACAGCTTTAGAGAAGATATATCGTGGCTTCTTC 1836
QY 121 TCGTGAGGACACATTTGCAAGCACAAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 180
Db 1837 TCGTGAGGACACATTTGCAAGCACAAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 1896
QY 181 CCAGTTTCATCTCATGAGGGGTACATGATGGATTGACAGCCCATCAGGCCGGGTGG 240
Db 1897 CCAGTTTCATCTCATGAGGGGTACATGATGGATTGACAGCCCATCAGGCCGGGTGG 1956
QY 241 TAATATTCTACAATTTGGGAAGTAAGCTGATGGAAACAGGAAATTTATCAGAGATGAAGA 300
Db 1957 TAATATTCTACAATTTGGGAAGTAAGCTGATGGAAACAGGAAATTTATCAGAGATGAAGA 2016

QY 301 AACTGAAGTACAAGAGCAGATGAATCTCTTAATTCAGATGGGAATGCCTCAGGGTAGC 360
Db 2017 AACTGAAGTACAAGAGCAGATGAATCTCTTAATTCAGATGGGAATGCCTCAGGGTAGC 2076
QY 361 TAGCATGGAACAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT 420
Db 2077 TAGCATGGAACAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACT 2136
QY 421 GAAAGAGTTGAATGACTGGCTAACCAAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 480
Db 2137 GAAAGAGTTGAATGACTGGCTAACCAAAACAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAA 2196
QY 481 GCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCAAAGTACAAACATAAAGTGGCTTCA 540
Db 2197 GCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCAAAGTACAAACATAAAGTGGCTTCA 2256
QY 541 AGAAGATCTAGAACAAACAAAGTCAAGGTCAATTTCTCTCACTCACATGCTGGTGGTAGT 600
Db 2257 AGAAGATCTAGAACAAACAAAGTCAAGGTCAATTTCTCTCACTCACATGCTGGTGGTAGT 2316
QY 601 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAAGAAACAACTTAAGGTATTGGG 660
Db 2317 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAAGAAACAACTTAAGGTATTGGG 2376
QY 661 AGATCGATGGGCAACACATCTCTAGATGGACAGAGACCGTGGTCTTTTACAAGA --- 717
Db 2377 AGATCGATGGGCAACACATCTCTAGATGGACAGAGACCGTGGTCTTTTACAAGACCA 2436
QY 718 ----- 717
Db 2437 GCCTGACCTAGCTCCTGGACTGACCACATATTGGAGCCTCTCTCTACTCAGACTGTTACTCT 2496
QY 718 ----- 717
Db 2497 GGTGACACAACCTGTGGTTACTAAGGAAACTGCCATCTCCAAACTAGAAATGCCATCTTC 2556
QY 718 ----- 717
Db 2557 CTTGATGTTGGAGGTACCTACTCATAGATTACTGCAACAGTTCCCTCCCTGGACCTGGAAAAA 2616
QY 718 ----- 717
Db 2617 GTTTCTTGGCTTACAGAAGCTGAAACAACTGCCAATGTCTACAGGATGCTACCCG 2676
QY 718 ----- 717
Db 2677 TAAGGAAAGGCTCCTAGAAAGACTCCAAGGGAGTAAAGAGCTGATGAAACAATGGCAAGA 2736
QY 718 ----- 717
Db 2737 CCTCCAAGGTGAAATTGAAGCTCACACAGATGTTTATCACAACTGGATGAAACACGCCA 2796
QY 718 ----- 717
Db 2797 AAAAATCCTGAGATCCCTGGAAGGTTCCGATGATGCAGTCTCTGTACAAAGACGTTTGA 2856
QY 718 ----- 717
Db 2857 TAACATGAACCTCAAGTGGAGTGAACCTCGGAAAAAGTCTCTCAACATTAGGTCCCATTT 2916
QY 718 -----CAGTTCTGACCAGTGGAAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGTG 771
Db 2917 GGAAGCCAGTTCTGACCAGTGGAAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGTGTG 2976
QY 772 GCTACAGCTGAAAGATGATGAATTAAAGCCGAGGACCTATTGGAGGCGACTTCCAGC 831
Db 2977 GCTACAGCTGAAAGATGATGAATTAAAGCCGAGGACCTATTGGAGGCGACTTCCAGC 3036
QY 832 AGTTTCAAGACAGAAACGATGTACATAGGCGCTTCAAGAGGGAATTGAAACATAAAGAACCC 891
Db 3037 AGTTTCAAGACAGAAACGATGTACATAGGCGCTTCAAGAGGGAATTGAAACATAAAGAACCC 3096

QY	892	TCTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGAAAGG	951
Db	3097	TGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTGGAAAGG	3156
QY	952	ACTAGAGAAAACCTCTACCAGGAGCCCCAGAGAGCTGCCTCCTGAGGAGAGAGCCCGAGAATGT	1011
Db	3157	ACTAGAGAAAACCTCTACCAGGAGCCCCAGAGAGCTGCCTCCTGAGGAGAGAGCCCGAGAATGT	3216
QY	1012	CACCTCGGCTTCTACGAAAAGCAGGCTGAGGAGGTCAATACTAGTGGGAAAAAATTGAACCT	1071
Db	3217	CACCTCGGCTTCTACGAAAAGCAGGCTGAGGAGGTCAATACTAGTGGGAAAAAATTGAACCT	3276
QY	1072	GCACTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGAGACTCCAGGAACTTCA	1131
Db	3277	GCACTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGAGACTCCAGGAACTTCA	3336
QY	1132	AGAGGCCACGGATGAGCTGSACCTCAAGCTGCGCCAAGCTGAGGTGATCAAGGGATCCTG	1191
Db	3337	AGAGGCCACGGATGAGCTGSACCTCAAGCTGCGCCAAGCTGAGGTGATCAAGGGATCCTG	3396
QY	1192	GCAGCCCGTGGCGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGC	1251
Db	3397	GCAGCCCGTGGCGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGC	3456
QY	1252	ACTTCGAGGAGAAAATTGCGCCTCTGAAAGAGAAACGTGAGCCACGTCAATGACCTTGCTCG	1311
Db	3457	ACTTCGAGGAGAAAATTGCGCCTCTGAAAGAGAAACGTGAGCCACGTCAATGACCTTGCTCG	3516
QY	1312	CCAGCTTACCACCTTTGGGCATTTCAGCTCTCACCGTATAACCTCAGCACTCTTGAAGACCT	1371
Db	3517	CCAGCTTACCACCTTTGGGCATTTCAGCTCTCACCGTATAACCTCAGCACTCTTGAAGACCT	3576
QY	1372	GAACACCAGATGGAAGCTTCTGCAGGTGGCGTCGAGSACCGAGTCAGGCAGCTGCATGA	1431
Db	3577	GAACACCAGATGGAAGCTTCTGCAGGTGGCGTCGAGSACCGAGTCAGGCAGCTGCATGA	3636
QY	1432	AGCCACACAGGACTTTGGTCCAGCATCTCAGCACTTCTTTCCACGTCTGTCTCCAGGGTCC	1491
Db	3637	AGCCACACAGGACTTTGGTCCAGCATCTCAGCACTTCTTTCCACGTCTGTCTCCAGGGTCC	3696
QY	1492	CTGGGAGAGAGCCATCTCGCAACAAAGTGCCCTACTATATCAACCACGAGACTCAAAAC	1551
Db	3697	CTGGGAGAGAGCCATCTCGCAACAAAGTGCCCTACTATATCAACCACGAGACTCAAAAC	3756
QY	1552	AACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAA	1611
Db	3757	AACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAA	3816
QY	1612	TGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAAGGCCCTTTG	1671
Db	3817	TGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAAGGCCCTTTG	3876
QY	1672	CTTGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGACCAGCACAACTCAAGCA	1731
Db	3877	CTTGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGACCAGCACAACTCAAGCA	3936
QY	1732	AAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTGACCACATTTATGACCG	1791
Db	3937	AAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTGACCACATTTATGACCG	3996
QY	1792	CCTGGAGCAAGAGCACAAATTTGGTCAACGTCCTCTCTCGTGGATATGTCTGAA	1851
Db	3997	CCTGGAGCAAGAGCACAAATTTGGTCAACGTCCTCTCTCGTGGATATGTCTGAA	4056
QY	1852	CTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCGTGTCTCTTTAA	1911
Db	4057	CTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCGTGTCTCTTTAA	4116
QY	1912	AACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAGATACCTTTTCAA	1971
Db	4117	AACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAGATACCTTTTCAA	4176
QY	1972	GCAAGTGGCAAGTTCAACAGGATTTGTGACCAGCGCAGGCTGGGCGCTCTCTGCAATGA	2031

Db	4177	GCAAGTGGCAAGTTCACAGGATTTGTGACCCAGCGCAGGCTGGGCTCTTCTGCATGA	4236
QY	2032	TTCTATCCAA	2041
Db	4237	TTCTATCCAA	4246
RESULT 12			
AAD37262			
ID	AAD37262 standard; DNA; 4990 BP.		
XX			
AC	AAD37262;		
XX			
DT	21-AUG-2002 (first entry)		
XX			
DE	Adeno-associated virus (AAV) vector plasmid, AAV-CMV-delta3990.		
XX			
KW	Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;		
KW	adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;		
KW	Becker muscular dystrophy; ds.		
XX			
OS	Chimeric - Homo sapiens.		
OS	Chimeric - Cytomegalovirus.		
OS	Chimeric - Unidentified.		
XX			
PN	WO200183695-A2.		
XX			
PD	08-NOV-2001.		
XX			
PF	27-APR-2001; 2001WO-US13677.		
XX			
PR	28-APR-2000; 2000US-200777P.		
XX			
PA	(XIAO/) XIAO X.		
XX			
PI	Xiao X;		
XX			
DR	WPI; 2002-049342/06.		
XX			
PT	New dystrophin minigene for treating Duchenne or Becker muscular		
PT	dystrophy comprises an N-terminal domain or modified N-terminal domain,		
PT	rod repeats, H1 and H4 domains and a cysteine rich domain of a		
PT	dystrophin gene -		
XX			
PS	Example 1; Page 67-68; 71pp; English.		
XX			

The present invention relates to an isolated nucleotide sequence encoding a dystrophin minigene. The minigene comprises N-terminal or modified N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control element. The dystrophin minigene in operable linkage with an expression control element, in a recombinant adeno-associated virus or retrovirus is useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is AAV vector plasmid construct containing human dystrophin minigenes, a cytomegalovirus (CMV) promoter and a small polyA signal sequence.

SQ Sequence 4990 BP; 1439 A; 1185 C; 1208 G; 1158 T; 0 other;

Query Match 75.6%; Score 1542; DB 24; Length 4990;
Best Local Similarity 80.7%; Pred. No. 0;
Matches 2041; Conservative 0; Mismatches 0; Indels 499

		Conservative	Mismatches	Indels	Gaps
Qy	1	TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAGTCATTGGCAGTTTCATTGATGGAGAG	0	0	0
Db	1741	TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAGTCATTGGCAGTTTCATTGATGGAGAG	0	0	0
Qy	61	TGAAGTAAACCTGGACCGTTATCAAACAGCTTTAGAAAGATATTATCGTGGCTTCTTTC	0	0	0
Db	1801	TGAAGTAAACCTGGACCGTTATCAAACAGCTTTAGAAAGATATTATCGTGGCTTCTTTC	0	0	0

QY 718 -----CAGTTCTGACCAGTGGAAAGCGTCTGCACCTTTCTCTCGAGGAACCTTCTGGTGTG 771
Db 2941 GGAAGCCAGTTCTGACCAGTGGAAAGCGTCTGCACCTTTCTCTCGAGGAACCTTCTGGTGTG 3000
QY 772 GCTACAGCTGAAAGATGATGAATTAAGCCGCGCAGGACCTATTGGAGGCGACTTTCCAGC 831
Db 3001 GCTACAGCTGAAAGATGATGAATTAAGCCGCGCAGGACCTATTGGAGGCGACTTTCCAGC 3060
QY 832 AGTTTCAGAACGACGATGTACATAGGGCCCTTCAAGAGGGAATTTGAAACCTAAAGAAC 891
Db 3061 AGTTTCAGAACGACGATGTACATAGGGCCCTTCAAGAGGGAATTTGAAACCTAAAGAAC 3120
QY 892 TGTAAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAAG 951
Db 3121 TGTAAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAAG 3180
QY 952 ACTAGAGAAACTCTACAGGAGCCCGAGAGCTGCCCTCTGAGGAGAGAGCCCGAGAAATGT 1011
Db 3181 ACTAGAGAAACTCTACAGGAGCCCGAGAGCTGCCCTCTGAGGAGAGAGCCCGAGAAATGT 3240
QY 1012 CACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACCTGAGGTGATCAAGGGATCCTG 1071
Db 3241 CACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACCTGAGGTGATCAAGGGATCCTG 3300
QY 1072 GCACCTCCGCTGACTGGCAGAGAAATAGATGAGACCCCTTGAAGAGACTCCAGGAACTTCA 1131
Db 3301 GCACCTCCGCTGACTGGCAGAGAAATAGATGAGACCCCTTGAAGAGACTCCAGGAACTTCA 3360
QY 1132 AGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATCCTG 1191
Db 3361 AGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATCCTG 3420
QY 1192 GCAGCCCGTGGCGGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGC 1251
Db 3421 GCAGCCCGTGGCGGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGC 3480
QY 1252 ACTTCGAGGAGAAATTTGCGCCTCTGAAAGAGAACCTGAGCCACGTCAATGACCTTGCTCG 1311
Db 3481 ACTTCGAGGAGAAATTTGCGCCTCTGAAAGAGAACCTGAGCCACGTCAATGACCTTGCTCG 3540
QY 1312 CCAGCTTACCCTTTGGGCTTTCAGCTCTCACCTGATCAACCTTCAGCAGCTCTGGAAGACCT 1371
Db 3541 CCAGCTTACCCTTTGGGCTTTCAGCTCTCACCTGATCAACCTTCAGCAGCTCTGGAAGACCT 3600
QY 1372 GAACACCCAGATGGAAGCTTCTGCAGGTGGCCGCTCGAGGACCGAGTCAAGGAGCTGCATGA 1431
Db 3601 GAACACCCAGATGGAAGCTTCTGCAGGTGGCCGCTCGAGGACCGAGTCAAGGAGCTGCATGA 3660
QY 1432 AGCCACACAGGAGCTTTGGTCCAGCATCTCAGCAGCTTTCTTTCCACGTCTGTCCAGGGTCC 1491
Db 3661 AGCCACACAGGAGCTTTGGTCCAGCATCTCAGCAGCTTTCTTTCCACGTCTGTCCAGGGTCC 3720
QY 1492 CTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACACGAGACTCAAAAC 1551
Db 3721 CTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACACGAGACTCAAAAC 3780
QY 1552 AACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAA 1611
Db 3781 AACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAATAA 3840
QY 1612 TGTGAGATCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAGGCCCTTTG 1671
Db 3841 TGTGAGATCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAGGCCCTTTG 3900
QY 1672 CTTGGATCTCTGAGCCTGTGAGCTGATGTGATGCCCTTGACCCAGCACCACTCAAGCA 1731
Db 3901 CTTGGATCTCTGAGCCTGTGAGCTGATGTGATGCCCTTGACCCAGCACCACTCAAGCA 3960
QY 1732 AAAATGACCCAGCCCATGGATATCCTGCAGATTAATAATGTTTGACCACTATTTATGACCG 1791
Db 3961 AAAATGACCCAGCCCATGGATATCCTGCAGATTAATAATGTTTGACCACTATTTATGACCG 4020
QY 1792 CCTGGAGCAAGAGCACAACTTTGGTCAACGTCCCTCTCTGCGTGGATATGTGTCTGAA 1851

QY 121 TGCTGAGGACACATTGCAAGCACACAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 180
Db 1861 TGCTGAGGACACATTGCAAGCACACAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGA 1920
QY 181 CCAGTTTCATACTCATGAGGGGTACATGATGGATTGACAGCCCATCAGGGCCGGGTGG 240
Db 1921 CCAGTTTCATACTCATGAGGGGTACATGATGGATTGACAGCCCATCAGGGCCGGGTGG 1980
QY 241 TAATATTCFACAAATTTGGGAAGTAAGCTGATTGGAACAGGAAATTTATCAGAAGATGAAGA 300
Db 1981 TAATATTCFACAAATTTGGGAAGTAAGCTGATTGGAACAGGAAATTTATCAGAAGATGAAGA 2040
QY 301 AACTGAAAGTACAGAGACGATGAATCTCCTAAATTCAGATGGGAATGCCCTCAGGGTAGC 360
Db 2041 AACTGAAAGTACAGAGACGATGAATCTCCTAAATTCAGATGGGAATGCCCTCAGGGTAGC 2100
QY 361 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTTTAATGGATCTCCAGAACTCAGAAACT 420
Db 2101 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTTTAATGGATCTCCAGAACTCAGAAACT 2160
QY 421 GAAAGAGTTGAATGACTGGCTAACAAGAACAGAAAGAACAGAAAGTAAAGGGAAGA 480
Db 2161 GAAAGAGTTGAATGACTGGCTAACAAGAACAGAAAGAACAGAAAGTAAAGGGAAGA 2220
QY 481 GCCTCTTGACCTGATCTTGAAGACCTAAACGCCCAAGTACAAACATAGAGTGCTTCA 540
Db 2221 GCCTCTTGACCTGATCTTGAAGACCTAAACGCCCAAGTACAAACATAGAGTGCTTCA 2280
QY 541 AGAAGATCTAGAACAAAGAACAAAGTCAAGGTCAATTTCTCTCACTCACATGGTGGTAGT 600
Db 2281 AGAAGATCTAGAACAAAGAACAAAGTCAAGGTCAATTTCTCTCACTCACATGGTGGTAGT 2340
QY 601 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGG 660
Db 2341 TGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAGAACAACTTAAGGTATTGGG 2400
QY 661 AGATCGATGGGCAAAACATCTGTAGATGGACAGAACCGCTGGGTTCTTTTACRAGA --- 717
Db 2401 AGATCGATGGGCAAAACATCTGTAGATGGACAGAACCGCTGGGTTCTTTTACRAGACCA 2460
QY 718 ----- 717
Db 2461 GCCTGACCTAGCTCCTGGACTGACCACCTATTGGAGCCTCTCCTACTCAGACTGTTACTCT 2520
QY 718 ----- 717
Db 2521 GGTGACACAACCTGTGGTTACTAAGGAAACTGCCATCTCCAACCTAGAAATGCCATCTTC 2580
QY 718 ----- 717
Db 2581 CTTGATGTTGGAGGTACCTACTCATAGATTACTGCAACAGTTCCCCCTGGACCTGGAAAA 2640
QY 718 ----- 717
Db 2641 GTTCTTGCCCTTACAGAAAGCTGAAACAACTGCCAATGCTCCTACAGGATGCTACCCG 2700
QY 718 ----- 717
Db 2701 TAAGGAAAGGCTCCTAGAGACTCCAAGGAGTAAAGAGCTGATGAACAAATGGCAAGA 2760
QY 718 ----- 717
Db 2761 CCTCCAGGTGAAATTTGAAGCTCACACAGATGTTTATCACAACCTGGATGAAACAGCCA 2820
QY 718 ----- 717
Db 2821 AAAAATCCTGATCCCTGGAAAGTTCCCGATGATGCAGTCTGTTTACAAAGACGTTTGA 2880
QY 718 ----- 717
Db 2881 TAACATGAACCTCAAGTGGAGTGAACCTTCGGAAAAAGTCTCTCAACATTTAGGTCCCAATT 2940

Db 4021 CCTGGAGCAAGAGCAACAATTTGGTCAACGTCCTCTCTGGTGGATATGTGTGAA 4080
QY 1852 CTGGCTGCTGAATCTTATGATACGGGACGAACAGGGAGGATCCGTCTCTTTAA 1911
Db 4081 CTGGCTGCTGAATCTTATGATACGGGACGAACAGGGAGGATCCGTCTCTTTAA 4140
QY 1912 AACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAGACAAAGTACAGATACCTTTCAA 1971
Db 4141 AACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAGACAAAGTACAGATACCTTTCAA 4200
QY 1972 GCAAGTGGCAAGTTCAACAGGATTTTGTGACCGCAGGCTGGGCCTCTCTGTCATGA 2031
Db 4201 GCAAGTGGCAAGTTCAACAGGATTTTGTGACCGCAGGCTGGGCCTCTCTGTCATGA 4260
QY 2032 TTCTATCCAA 2041
Db 4261 TTCTATCCAA 4270

RESULT 13
AAD37230
ID AAD37230 standard; DNA; 4182 BP.
XX AC AAD37230;
XX 21-AUG-2002 (first entry)
DE Human dystrophin minigene delta4173.

XX Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX Homo sapiens.

OS WO200183695-A2.
XX 08-NOV-2001.
PD 27-APR-2001; 2001WO-US13677.
XX 28-APR-2000; 2000US-200777P.
PR (XIAO/) XIAO X.
XX Xiao X;
XX WPI; 2002-049342/06.

New dystrophin minigene for treating Duchenne or Becker muscular dystrophy comprises an N-terminal domain or modified N-terminal domain, rod repeats, H1 and H4 domains and a cysteine rich domain of a dystrophin gene -
Example 1; Page 43-44; 71pp; English.

The present invention relates to an isolated nucleotide sequence encoding a dystrophin minigene. The minigene comprises N-terminal or modified N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control element. The dystrophin minigene in operable linkage with an expression control element, in a recombinant adeno-associated virus or retrovirus is useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is human dystrophin minigene delta4173 containing nucleotides 1-1992 (N-terminus, hinge H1 and rods R1, R2 and R3), 8059-10227 (rods R22, R23 and R24, hinge H4 and CR domain) and 11047-11058 (dystrophin last 3 amino acids).

Sequence 4182 BP; 1309 A; 927 C; 970 G; 976 T; 0 other;

Query Match 66.6%; Score 1359; DB 24; Length 4182;
Best Local Similarity 75.2%; Pred. No. 0;
Matches 2041; Conservative 0; Mismatches 0; Indels 672; Gaps 1;
QY 1 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAGTCAATTTGGCAGTTCATTGATGGAGAG 60
Db 960 TCCTTCACAGCATTTGGAAGCTCCTGAAGACAAGTCAATTTGGCAGTTCATTGATGGAGAG 1019
QY 61 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAGAAGTATTATCGTGGCTTCTTTC 120
Db 1020 TGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAGAAGTATTATCGTGGCTTCTTTC 1079
QY 121 TGCTGAGGACACATTGCAAGCACACAAGGAGAGATTTCTTAATGATGTGGAAGTGGTAAAGA 180
Db 1080 TGCTGAGGACACATTGCAAGCACACAAGGAGAGATTTCTTAATGATGTGGAAGTGGTAAAGA 1139
QY 181 CCAGTTTCATACTCATGAGGGGTACATGATGATTTGACAGCCCATCAGGGCCGGTTGG 240
Db 1140 CCAGTTTCATACTCATGAGGGGTACATGATGATTTGACAGCCCATCAGGGCCGGTTGG 1199
QY 241 TAAATATTCTACAATTGGGAAGTAAGCTGATTTGGAACAGGAAAATTTATCAGAAGATGAAGA 300
Db 1200 TAAATATTCTACAATTGGGAAGTAAGCTGATTTGGAACAGGAAAATTTATCAGAAGATGAAGA 1259
QY 301 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTTCAAGATGGGAATGCCTCAGGGTAGC 360
Db 1260 AACTGAAGTACAAGAGCAGATGAATCTCCTAAATTTCAAGATGGGAATGCCTCAGGGTAGC 1319
QY 361 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAAACT 420
Db 1320 TAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAATCAGAAACT 1379
QY 421 GAAAGAGTTGAATGACTGGCTAAACAAAAACAGAAGAAAGAAACAAAGGAAATGGAGGAAGA 480
Db 1380 GAAAGAGTTGAATGACTGGCTAAACAAAAACAGAAGAAAGAAACAAAGGAAATGGAGGAAGA 1439
QY 481 GCCTCTTGACCTGATCTTGAAGACCTTAAACGCCCAAGTACAACAACATAGGTGCTTCA 540
Db 1440 GCCTCTTGACCTGATCTTGAAGACCTTAAACGCCCAAGTACAACAACATAGGTGCTTCA 1499
QY 541 AGAAGATCTAGAACACAAGAACAGTCAAGGTCAATTTCTCTCACTCACAATGGTGGTAGT 600
Db 1500 AGAAGATCTAGAACACAAGAACAGTCAAGGTCAATTTCTCTCACTCACAATGGTGGTAGT 1559
QY 601 TGATGAATCTAGTGGAGATCAGCAACTGCTGCTTTGGAAGAAACAACATTAAGGTATTTGGG 660
Db 1560 TGATGAATCTAGTGGAGATCAGCAACTGCTGCTTTGGAAGAAACAACATTAAGGTATTTGGG 1619
QY 661 AGATCGATGGGCAACATCTGTAGATGGACAGAAGACCGCTGGGTCTTTTACAAGA--- 717
Db 1620 AGATCGATGGGCAACATCTGTAGATGGACAGAAGACCGCTGGGTCTTTTACAAGACAT 1679
QY 718 ----- 717
Db 1680 CCTTCTCAAAATGGCAACGCTTACTGAAGAACAGTGCCTTTTATAGTGCATGGCTTTCAGA 1739
QY 718 ----- 717
Db 1740 AAAAGAAGATGCAGTGAACAAGATTTCACACAACACTGGCTTTAAAGATCAAAATGAAATGTT 1799
QY 718 ----- 717
Db 1800 ATCAAGTCTTCAAAAAACTGGCCGTTTTTAAAGCGGATCTAGAAAAGAAAAGCAATCCAT 1859
QY 718 ----- 717
Db 1860 GGGCAAACTGTATTCACTCAACAAGATCTTCTTTCAACACTGAAGAATAAGTCAGTGAC 1919
QY 718 ----- 717
Db 1920 CCAGAAGACGGAAGCATGGCTGGATAACTTTGCCCGGTGTTGGGATAATTTAGTCCAAA 1979
QY 718 ----- 717


```

Db 1980 ACTTGAAGAGGTACAGCACAGACTCATAGATTACTGCAACAGTTTCCCCCTGGACCTGGA 2039
QY 718 ----- 717
Db 2040 AAAGTTTCTTGCTGGCTTACAGAAGCTGAAACAACTGCCAATGTCTACAGGATGCTAC 2099
QY 718 ----- 717
Db 2100 CCGTAAGGAAGGCTCCTAGAAGACTCCAAAGGAGTAAAGAGCTGTATGAAACAATGGCA 2159
QY 718 ----- 717
Db 2160 AGACCTCCAAGGTGAAATTTGAAGCTCACACAGATGTTTATCACAAACCTGGATGAAACAG 2219
QY 718 ----- 717
Db 2220 CCAAAAAATCCTGAGATCCCTGGAGGTTCCGATGATGCAGTCCCTGTTACAAAGACGTTT 2279
QY 718 ----- 717
Db 2280 GGATAACATGAACCTCAAGTGGAGTGAACCTTCGGAAGAAAGTCTCTCAACATTAGGTCCCA 2339
QY 718 -----CAGTTCTGACCAAGTGAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGT 768
Db 2340 TTTGGAAGCCAGTTCTGACCAAGTGAAGCGTCTGCACCTTTCTCTGCAGGAACCTTCTGGT 2399
QY 769 GTGGCTACAGCTGAAAGATGATGAATTAAGCCGCGCAGGCACCTATTGGAGGCGACTTTCC 828
Db 2400 GTGGCTACAGCTGAAAGATGATGAATTAAGCCGCGCAGGCACCTATTGGAGGCGACTTTCC 2459
QY 829 AGCAGTTTCAAGAGCAGAACCATGTACATAGGGCCTTCAAGAGGGAATTTGAAACTAAAGA 888
Db 2460 AGCAGTTTCAAGAGCAGAACCATGTACATAGGGCCTTCAAGAGGGAATTTGAAACTAAAGA 2519
QY 889 ACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTTGA 948
Db 2520 ACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCTTTTGA 2579
QY 949 AGGACTAGAGAAACTCTACCAGGAGCCAGAGAGCTGCCTCCTGAGGAGAGAGCCAGAA 1008
Db 2580 AGGACTAGAGAAACTCTACCAGGAGCCAGAGAGCTGCCTCCTGAGGAGAGAGCCAGAA 2639
QY 1009 TGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATATCTAGTGGGAAAAATTGAA 1068
Db 2640 TGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATATCTAGTGGGAAAAATTGAA 2699
QY 1069 CCTGCACTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGACTCCAGGAAT 1128
Db 2700 CCTGCACTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGACTCCAGGAAT 2759
QY 1129 TCAAGAGCCACGGATGAGCTGGACCTCAAGCTCGCCAAAGCTGAGGTGATCAAGGGATC 1188
Db 2760 TCAAGAGCCACGGATGAGCTGGACCTCAAGCTCGCCAAAGCTGAGGTGATCAAGGGATC 2819
QY 1189 CTGGCAGCCCGTGGCGGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGAGAAAGTCAA 1248
Db 2820 CTGGCAGCCCGTGGCGGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGAGAAAGTCAA 2879
QY 1249 GGCACCTTCAGAGAGAAATTTGGCCCTCTGAAAGAGAAACGTGAGCCACGTCAATGACCTTGC 1308
Db 2880 GGCACCTTCAGAGAGAAATTTGGCCCTCTGAAAGAGAAACGTGAGCCACGTCAATGACCTTGC 2939
QY 1309 TCGCCAGCTTACCACCTTTGGGCATTTAGCTCTCAGCTCTCAGCTTAAACCTCAGCAGCTTGAAGA 1368
Db 2940 TCGCCAGCTTACCACCTTTGGGCATTTAGCTCTCAGCTCTCAGCTTAAACCTCAGCAGCTTGAAGA 2999
QY 1369 CCTGAACACCAGATGGAAGCTTCTGCAGGTGGCCGCTGAGGACCGAGTCAAGGAGCTGCA 1428
Db 3000 CCTGAACACCAGATGGAAGCTTCTGCAGGTGGCCGCTGAGGACCGAGTCAAGGAGCTGCA 3059
QY 1429 TGAAGCCACAGGAGCTTTGGTCCAGCATCTCAGCAGCTTTCTTTCCACGCTCTGTCCAGGG 1488

```

```

Db 3060 TGAAGCCACAGGGACTTTGGTCCAGCATCTCAGACATTTCTTTCCAGTCTGTCCAGGG 3119
QY 1489 TCCCTGGGAGAGAGCCCATCTCGCCAAACAAAGTGCCTACTATATCAACACGAGACTCA 1548
Db 3120 TCCCTGGGAGAGAGCCCATCTCGCCAAACAAAGTGCCTACTATATCAACACGAGACTCA 3179
QY 1549 AACAACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAA 1608
Db 3180 AACAACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGAA 3239
QY 1609 TAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAGGCCCCT 1668
Db 3240 TAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAGGCCCCT 3299
QY 1669 TTGCTTGGATCTCTTGAAGCTGTGAGCTGTGATGCCTTGGACGAGCACAACCTCAA 1728
Db 3300 TTGCTTGGATCTCTTGAAGCTGTGAGCTGTGATGCCTTGGACGAGCACAACCTCAA 3359
QY 1729 GCAAAATGACGAGCCCATGGATATCCTGCAGATTATTAATGTTTGACCACTATTTATGA 1788
Db 3360 GCAAAATGACGAGCCCATGGATATCCTGCAGATTATTAATGTTTGACCACTATTTATGA 3419
QY 1789 CCGCCTGGAGCAAGAGCACAACAAATTTGGTCAACGTCCTCTCTGCGTGGATATGTGTCT 1848
Db 3420 CCGCCTGGAGCAAGAGCACAACAAATTTGGTCAACGTCCTCTCTGCGTGGATATGTGTCT 3479
QY 1849 GAACTGGCTGCTGAATGTTTATGATACGGGACGAACAGGAGGATCCGTGTCTCTTT 1908
Db 3480 GAACTGGCTGCTGAATGTTTATGATACGGGACGAACAGGAGGATCCGTGTCTCTTT 3539
QY 1909 TAAAACTGGCATCATTTCCCTGTGTAAGACACATTTGGAGAGACAAGTACAGATACCTTTT 1968
Db 3540 TAAAACTGGCATCATTTCCCTGTGTAAGACACATTTGGAGAGACAAGTACAGATACCTTTT 3599
QY 1969 CAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAAGCGAGGCTGGGCTCTTCTGCA 2028
Db 3600 CAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAAGCGAGGCTGGGCTCTTCTGCA 3659
QY 2029 TGATTCTATCCAA 2041
Db 3660 TGATTCTATCCAA 3672

```

```

RESULT 14
AAD37232
ID AAD37232 standard; DNA; 2169 BP.
XX
AC AAD37232;
XX
DT 21-AUG-2002 (first entry)
XX
DE Human dystrophin rod, hinge and CR domain regions encoding DNA #1.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Homo sapiens.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
PI XIAO X;
XX
DR WPI; 2002-049342/06.
XX

```


PT New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -

XX Example 1; Page 45-46; 71pp; English.

XX The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) or Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is human
CC dystrophin DNA fragment encoding rods R22, R23 and R24, hinge H4 and
CC CR domain regions.

XX Sequence 2169 BP; 623 A; 529 C; 524 G; 493 T; 0 other;

Query Match 65.1%; Score 1328.8; DB 24; Length 2169;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1333; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY	702	GGGTTCTTTTACAAGACAGTTCTGACCAGTGAAGAGATGATGAATTAAGCCGGCAGGCACCTATTGGAGGCG	761
Db	333	GGTCCCATTTGGAAGCCAGTTCTTGACCAGTGAAGCGTCTGCACCTTTCTCTGCAGGAAC	392
QY	762	TTCTGGTGTGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTGGAGGCG	821
Db	393	TTCTGGTGTGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTGGAGGCG	452
QY	822	ACTTTCAGCAGTTTCAGAAAGCAGACGATGTACATAGGCGCTTCAAGAGGGAATGAAAA	881
Db	453	ACTTTCAGCAGTTTCAGAAAGCAGACGATGTACATAGGCGCTTCAAGAGGGAATGAAAA	512
QY	882	CTAAAGAACCTGTAAATCATGAGTACTCTTGGAGACTGTACGAATATTTCTGACAGAGCAGC	941
Db	513	CTAAAGAACCTGTAAATCATGAGTACTCTTGGAGACTGTACGAATATTTCTGACAGAGCAGC	572
QY	942	CTTTGGAAGGACTAGAGAACTCTACCAGGAGCCAGAGAGCTGCCTCCTGAGGAGAGAG	1001
Db	573	CTTTGGAAGGACTAGAGAACTCTACCAGGAGCCAGAGAGCTGCCTCCTGAGGAGAGAG	632
QY	1002	CCCAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGAAA	1061
Db	633	CCCAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGAAA	692
QY	1062	AATTGAACCTGCACCTCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAAGACTCC	1121
Db	693	AATTGAACCTGCACCTCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAAGACTCC	752
QY	1122	AGGAACCTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGATCA	1181
Db	753	AGGAACCTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGATCA	812
QY	1182	AGGGATCCTGGCAGCCCGTGGCGGATCTCCTCATTTGACTCTCTCCAAAGATCACCTCGAGA	1241
Db	813	AGGGATCCTGGCAGCCCGTGGCGGATCTCCTCATTTGACTCTCTCCAAAGATCACCTCGAGA	872
QY	1242	AAGTCAAGGCACCTTCGAGGAGAAAAATTGCGCCTCTGAAAGAGAAACGTGAGCCACGTCAATG	1301
Db	873	AAGTCAAGGCACCTTCGAGGAGAAAAATTGCGCCTCTGAAAGAGAAACGTGAGCCACGTCAATG	932
QY	1302	ACCTTGCTGCCAGCTTACCACCTTTGGGCAATTCAGCTCTCACCCGTATAACCTCAGCACTC	1361
Db	933	ACCTTGCTGCCAGCTTACCACCTTTGGGCAATTCAGCTCTCACCCGTATAACCTCAGCACTC	992
QY	1362	TGGAAGACCTGAACACACAGATGGAAGCTTCTGACGGTGGCCGTCGAGGACCGAGTCAGGC	1421
Db	993	TGGAAGACCTGAACACACAGATGGAAGCTTCTGACGGTGGCCGTCGAGGACCGAGTCAGGC	1052

QY	1422	AGCTGCATGAAGCCACACAGGACTTTGGTCCAGCATCTCAGCAGCTTTCTTTCCACGTCTG	1481
Db	1053	AGCTGCATGAAGCCACACAGGACTTTGGTCCAGCATCTCAGCAGCTTTCTTTCCACGTCTG	1112
QY	1482	TCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCACG	1541
Db	1113	TCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCACG	1172
QY	1542	AGACTCAAAACAACTTGTCTGGGACCATCCCAAAATGACAGAGTCTACCACTCTTTAGCTG	1601
Db	1173	AGACTCAAAACAACTTGTCTGGGACCATCCCAAAATGACAGAGTCTACCACTCTTTAGCTG	1232
QY	1602	ACCTGAATAATGTCTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGA	1661
Db	1233	ACCTGAATAATGTCTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGA	1292
QY	1662	AGGCCCTTTGCTGGATCTCTTGAGCCTGTCTCAGCTGCATGTGATGCCTTGGACCAGCACA	1721
Db	1293	AGGCCCTTTGCTGGATCTCTTGAGCCTGTCTCAGCTGCATGTGATGCCTTGGACCAGCACA	1352
QY	1722	ACCTCAAGCAAAATGACCAGCCCATGGATATCTCTGCAGATTATTAATTGTTGACCACCTA	1781
Db	1353	ACCTCAAGCAAAATGACCAGCCCATGGATATCTCTGCAGATTATTAATTGTTGACCACCTA	1412
QY	1782	TTTATGACCGCCTGGAGCAAGAGCACAACAATTTGGTCAACGTCCTCTCTGCGTGGATA	1841
Db	1413	TTTATGACCGCCTGGAGCAAGAGCACAACAATTTGGTCAACGTCCTCTCTGCGTGGATA	1472
QY	1842	TGTGTCTGAACTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCGTGTCC	1901
Db	1473	TGTGTCTGAACTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCGTGTCC	1532
QY	1902	TGTCTTTTAAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAGAT	1961
Db	1533	TGTCTTTTAAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAGAT	1592
QY	1962	ACCTTTTCAAGCAAGTGGCAAGTTCACACAGGATTTTGTGACCAGCGAGGCTGGGCTCC	2021
Db	1593	ACCTTTTCAAGCAAGTGGCAAGTTCACACAGGATTTTGTGACCAGCGAGGCTGGGCTCC	1652
QY	2022	TTCTGCATGATTCTATCCAA	2041
Db	1653	TTCTGCATGATTCTATCCAA	1672

RESULT 15

AAD37241

ID AAD37241 standard; DNA; 1821 BP.

XX AAD37241;

AC AAD37241;

XX 21-AUG-2002 (first entry)

DT Human dystrophin rod, hinge and CR domain regions encoding DNA #2.

XX Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;

DE adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;

XX Becker muscular dystrophy; ds.

XX Homo sapiens.

OS WO200183695-A2.

PN 08-NOV-2001.

XX 27-APR-2001; 2001WO-US13677.

XX 28-APR-2000; 2000US-200777P.

PR (XIAO/) XIAO X.

XX XIAO X;

PI

XX WPI; 2002-049342/06.
DR
XX
PT New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -
PT
XX
PS Example 1; Page 52-53; 71pp; English.
XX
CC The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is human
CC dystrophin DNA fragment encoding rods R23 and R24, hinge H4 and CR
CC domain regions.
XX
SQ Sequence 1821 BP; 506 A; 451 C; 447 G; 417 T; 0 other;
Query Match 64.9%; Score 1324; DB 24; Length 1821;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1324; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 718 CAGTTCTGACGAGTGGGAGCGTCTGCACCTTTCTCTGCAGGAACTTCTGGTGCTACA 777
Db 1 CAGTTCTGACGAGTGGGAGCGTCTGCACCTTTCTCTGCAGGAACTTCTGGTGCTACA 60
QY 778 GCTGAAAGATGATGAATTAAGCCGCGAGGCACCTATTGGAGGCGACTTCTCCAGCAGTTCA 837
Db 61 GCTGAAAGATGATGAATTAAGCCGCGAGGCACCTATTGGAGGCGACTTCTCCAGCAGTTCA 120
QY 838 GAAGCAGAACGATGTACATAGGGCCCTCAAGAGGGAATTTGAAACTAAAGAACCTGTAAT 897
Db 121 GAAGCAGAACGATGTACATAGGGCCCTCAAGAGGGAATTTGAAACTAAAGAACCTGTAAT 180
QY 898 CATGAGTACTCTTGAGACTGTACGAATATTCTGACAGAGCAGCCCTTTGGAAGGACTAGA 957
Db 181 CATGAGTACTCTTGAGACTGTACGAATATTCTGACAGAGCAGCCCTTTGGAAGGACTAGA 240
QY 958 GAACTCTACGAAAGCAGGCTGAGGAGGTCAATAGTGGGAAAAATTTGAACCTGCACTC 1017
Db 241 GAACTCTACGAAAGCAGGCTGAGGAGGTCAATAGTGGGAAAAATTTGAACCTGCACTC 300
QY 1018 GCTTCTACGAAAGCAGGCTGAGGAGGTCAATAGTGGGAAAAATTTGAACCTGCACTC 1077
Db 301 GCTTCTACGAAAGCAGGCTGAGGAGGTCAATAGTGGGAAAAATTTGAACCTGCACTC 360
QY 1078 CGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGAACTCCAGGAACTTCAAGAGGC 1137
Db 361 CGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGAACTCCAGGAACTTCAAGAGGC 420
QY 1138 CACGGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATCCTGGCAGCC 1197
Db 421 CACGGATGAGCTGGACCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATCCTGGCAGCC 480
QY 1198 CGTGGGCGATCTCCTCATTTGACTCTCTCCAAGATACCTCGAGAAAGTCAAGGCACCTCG 1257
Db 481 CGTGGGCGATCTCCTCATTTGACTCTCTCCAAGATACCTCGAGAAAGTCAAGGCACCTCG 540
QY 1258 AGGAGAAATTTGGCCCTCTGAAAGAGAACGCTGAGCCACGTCGAATGACCTTGCTCGCCAGCT 1317
Db 541 AGGAGAAATTTGGCCCTCTGAAAGAGAACGCTGAGCCACGTCGAATGACCTTGCTCGCCAGCT 600
QY 1318 TACCACCTTTGGGCAATTCAGCTCTACCGGTATACCTCAGCAGCTCTGGAAGACCTGAACAC 1377
Db 601 TACCACCTTTGGGCAATTCAGCTCTACCGGTATACCTCAGCAGCTCTGGAAGACCTGAACAC 660

QY 1378 CAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAAGGCAGCTGCATGAAGCCCA 1437
Db 661 CAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAAGGCAGCTGCATGAAGCCCA 720
QY 1438 CAGGACTTTGGTCCAGCATCTCAGCACCTTTCTTTCCACGCTCTGTCCAGGGTCCCTGGGA 1497
Db 721 CAGGACTTTGGTCCAGCATCTCAGCACCTTTCTTTCCACGCTCTGTCCAGGGTCCCTGGGA 780
QY 1498 GAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCACGAGACTCAAAACAACCTG 1557
Db 781 GAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCACGAGACTCAAAACAACCTG 840
QY 1558 CTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTAGCTGACCTGAATAATGTGAG 1617
Db 841 CTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTAGCTGACCTGAATAATGTGAG 900
QY 1618 ATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAAGGCCCTTTGCTTGA 1677
Db 901 ATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAAGGCCCTTTGCTTGA 960
QY 1678 TCTCTTGAGCCTGTGAGCTGATGATGCCTTGGACACGACCAACCTCAAGCAAAATGA 1737
Db 961 TCTCTTGAGCCTGTGAGCTGATGATGCCTTGGACACGACCAACCTCAAGCAAAATGA 1020
QY 1738 CCAGCCCATGGATATCCTGCGAGATTAATTTGTTGACCACTATTTATGACCCCTGGA 1797
Db 1021 CCAGCCCATGGATATCCTGCGAGATTAATTTGTTGACCACTATTTATGACCCCTGGA 1080
QY 1798 GCAAGAGCACAAACAATTTGGTCAACGTCCTCTCTGCGTGGATATGTGTGAACTGGCT 1857
Db 1081 GCAAGAGCACAAACAATTTGGTCAACGTCCTCTCTGCGTGGATATGTGTGAACTGGCT 1140
QY 1858 GCTGAATGTTTATGATACGGGACGAAACAGGAGGATCCGTGCTCTCTTTTAAACTGG 1917
Db 1141 GCTGAATGTTTATGATACGGGACGAAACAGGAGGATCCGTGCTCTCTTTTAAACTGG 1200
QY 1918 CATCATTTCCCTGTGTAAGCACATTTGGAAGACAAAGTACAGATACCTTTTCAAGCAAGT 1977
Db 1201 CATCATTTCCCTGTGTAAGCACATTTGGAAGACAAAGTACAGATACCTTTTCAAGCAAGT 1260
QY 1978 GGCAAGTTCAACAGGATTTTGTGACCAAGCGCAGGCTGGGCCCTCCTTCTGATGATCTAT 2037
Db 1261 GGCAAGTTCAACAGGATTTTGTGACCAAGCGCAGGCTGGGCCCTCCTTCTGATGATCTAT 1320
QY 2038 CCAA 2041
Db 1321 CCAA 1324

Search completed: April 25, 2003, 08:45:30
Job time : 379.763 secs

GenCore version 5.1.4.p5.4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 07:58:12 ; Search time 1925.12 Seconds
(without alignments)
17170.339 Million cell updates/sec

Title: US-09-845-416-12_COPY_960_3000
Perfect score: 2041
Sequence: 1 tccttcacagcatttggaag.....ttctgcatgattctatccaa 2041

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308016

Minimum DB seq length: 0
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: gb_gss:*
- 18: em_gss_hum:*
- 19: em_gss_inv:*
- 20: em_gss_pln:*
- 21: em_gss_vrt:*
- 22: em_gss_fun:*
- 23: em_gss_mam:*
- 24: em_gss_mus:*
- 25: em_gss_other:*
- 26: em_gss_pro:*
- 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	505	24.7	757	9 AL556247	AL556247 AL556247
2	494	24.2	620	14 BQ640063	BQ640063 he23g04.y
3	471.4	23.1	770	12 BG719710	BG719710 602690430
4	454.6	22.3	1490	11 BC009242	BC009242 Homo sapi
5	453.8	22.2	652	10 BB629984	BB629984 BB629984
6	453.8	22.2	2135	11 AK013510	AK013510 Mus muscu

7	449	22.0	704	10	BB610411	BB610411
8	428.6	21.0	854	9	AI196693	AI196693 ui53e10.y
9	410.4	20.1	599	10	BB666688	BB666688 BB666688
10	381.8	18.7	423	9	AA460476	AA460476 zx61e10.r
11	380.2	18.6	684	9	AL641565	AL641565 AL641565
12	374.8	18.4	898	12	BF182065	BF182065 601804604
13	368.6	18.1	633	9	AL796733	AL796733 AL796733
14	330	16.2	608	9	AL672616	AL672616 AL672616
15	329	16.1	532	10	BE334408	BE334408 ug97g06.y
16	325.6	16.0	532	13	BJ036615	BJ036615 BJ036615
17	315.4	15.5	541	13	BI546771	BI546771 603191346
18	310	15.2	488	13	BM488542	BM488542 pgm2n.pk0
19	307.4	15.1	834	13	BI729851	BI729851 603349511
20	303	14.8	579	9	AL121550	AL121550 DKFZp762L
21	297.8	14.6	597	9	AL795652	AL795652 AL795652
22	297.6	14.4	665	10	BE201973	BE201973 fl03h06.x
23	293.8	14.4	696	13	BJ075057	BJ075057 BJ075057
24	292.8	14.3	663	13	BM488464	BM488464 pgm2n.pk0
25	276.6	13.6	647	9	AL652775	AL652775 AL652775
26	265.8	13.0	784	12	BG212445	BG212445 RST32032
27	264.2	12.9	2334	11	BC011062	BC011062 Mus muscu
28	263.8	12.9	750	13	BI730168	BI730168 603349711
29	255.4	12.5	772	13	BI250598	BI250598 602993659
30	254.4	12.5	573	13	BI446561	BI446561 de24a10.y
31	250	12.2	502	9	AL602076	AL602076 DKFZp313B
32	246	12.1	835	13	BI553820	BI553820 603190772
33	225.8	11.1	733	12	BF144271	BF144271 601786885
34	211.4	10.4	501	9	AA771140	AA771140 vt18b05.r
35	206.4	10.1	479	14	BQ304046	BQ304046 QV2-BT063
36	199.6	9.8	726	10	BB627285	BB627285 BB627285
37	191.2	9.4	417	12	BF387048	BF387048 UI-R-CA1-
38	190.8	9.3	230	14	N87737	N87737 LL3635F Hum
39	184.8	9.1	307	12	BF226333	BF226333 uz48g06.y
40	184.6	9.0	475	12	BG207912	BG207912 RST27400
41	168.6	8.3	646	13	BI289102	BI289102 UI-R-DK0-
42	153.4	7.5	481	10	BE557463	BE557463 fl03h06.y
43	150.6	7.4	515	17	AZ780914	AZ780914 2M0018010
44	142.4	7.0	1002	12	BE798394	BE798394 601583022
45	142.2	7.0	426	14	BQ319056	BQ319056 IL5-CT051

ALIGNMENTS

RESULT 1
AL556247
LOCUS
DEFINITION AL556247 LTI_NFL006_PL2 Homo sapiens CDNA clone CS0DK001YB17 5
prime, mRNA sequence.
ACCESSION AL556247
VERSION AL556247.1 GI:12898746
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 757)
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
Location/Qualifiers
1. .757
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DK001YB17"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com

BASE COUNT 190 a 197 c 188 g 182 t
ORIGIN
Query Match 24.7%; Score 505; DB 9; Length 757;
Best Local Similarity 100.0%; Pred. No. 4.1e-112;
Matches 505; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1537 CCACGAGACTCAAAACAACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAAGTCTTT 1596
Db 128 CCACGAGACTCAAAACAACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAAGTCTTT 187
QY 1597 AGCTGACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT 1656
Db 188 AGCTGACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT 247
QY 1657 GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCCTGTCAGCTGCATGTGATGCCCTTGGACCA 1716
Db 248 GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCCTGTCAGCTGCATGTGATGCCCTTGGACCA 307
QY 1717 GCACAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAAATTTGTTGAC 1776
Db 308 GCACAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAAATTTGTTGAC 367
QY 1777 CACTATTATGACCGCCTGGAGCAAGAGCACAACTTTGGTCAACGTCCTCTCTGCGT 1836
Db 368 CACTATTATGACCGCCTGGAGCAAGAGCACAACTTTGGTCAACGTCCTCTCTGCGT 427
QY 1837 GGATATGTGCTGAAGTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCG 1896
Db 428 GGATATGTGCTGAAGTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCG 487
QY 1897 TGTCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTA 1956
Db 488 TGTCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTA 547
QY 1957 CAGATACCTTTTCAAGCAAGTGGCAAGTTTCAACAGGATTTTGTGACCAAGCGCAGGCTGG 2016
Db 548 CAGATACCTTTTCAAGCAAGTGGCAAGTTTCAACAGGATTTTGTGACCAAGCGCAGGCTGG 607
QY 2017 CCTCCTTCTGCATGATTCTATCCAA 2041
Db 608 CCTCCTTCTGCATGATTCTATCCAA 632

RESULT 2
BQ640063 620 bp mRNA linear EST 15-JUL-2002
LOCUS he23g04.y1 Human Retina cDNA (Un-normalized, unamplified): hd/he
DEFINITION Homo sapiens cDNA clone he23g04 5', mRNA sequence.
ACCESSION BQ640063
VERSION BQ640063.1 GI:21764522
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 620)
AUTHORS Wistow,G., Bernstein,S.L., Wyatt,M.K., Ray,S., Behal,A., Touchman
,J.W., Bouffard,G., Smith,D. and Peterson,K.
TITLE Expressed sequence tag analysis of human retina for the NEIBank
project: rebindin, an abundant, novel retinal cDNA and alternative
splicing of other retina-preferred gene transcripts
JOURNAL Mol. Vis. 8 (4), (2002) in press
COMMENT Contact: Wistow G
Section on Molecular Structure and Function

National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452
Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 23 row: g column: 04
Seq primer: M13RPI reverse primer (ABI).
Location/Qualifiers

FEATURES
source

1..620
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="he23g04"
): hd/he"
/tissue_type="Retina"
/dev_stage="Adult"
/lab_host="EMDH10B"
/note="Organ: Eye; Vector: pSPORT1; Neural retina tissue
was dissected from two 80 year old donors with no observed
eye disease. 100ug of total RNA was used for library in the
construction. A directionally cloned cDNA library in the
pSPORT1 vector (Life Technologies) was essentially following
Bioserve Biotechnology (Laurel MD) essentially following
the protocols of the SuperScript Plasmid System full
details of which are contained in the manufacturer's
instruction manual (http://www.lifetech.com/). First
strand synthesis was carried out using a Not I
primer-adaptor [5'-pGACTAGTTCTAGATCGGAGCGGCCGCC(T)15-3'
]. EST analysis was performed on the unamplified library
at the NIH Intramural Sequencing Center (NISC)."
BASE COUNT 165 a 148 c 144 g 163 t
ORIGIN

Query Match 24.2%; Score 494; DB 14; Length 620;
Best Local Similarity 100.0%; Pred. No. 1.8e-109;
Matches 494; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1548 AAACAACCTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAAGTCTTTAGCTGACCTGA 1607
Db 1 AAACAACCTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAAGTCTTTAGCTGACCTGA 60
QY 1608 ATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAGGCC 1667
Db 61 ATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAGGCC 120
QY 1668 TTTGCTTGGATCTCTTGAGCCCTGTCAGCTGCATGTGATGCCCTTGGACCAAGCAACCTCA 1727
Db 121 TTTGCTTGGATCTCTTGAGCCCTGTCAGCTGCATGTGATGCCCTTGGACCAAGCAACCTCA 180
QY 1728 AGCAAAATGACCAAGCCCATGGATATCCTGCAGATTATTAAATGTTTGACCACTATTATG 1787
Db 181 AGCAAAATGACCAAGCCCATGGATATCCTGCAGATTATTAAATGTTTGACCACTATTATG 240
QY 1788 ACCGCTGGAGCAAGAGACACAATTTGGTCAACGTCCTCTCTCGTGGATATGTGTC 1847
Db 241 ACCGCTGGAGCAAGAGACACAATTTGGTCAACGTCCTCTCTCGTGGATATGTGTC 300
QY 1848 TGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGAGGATCCCGTGTCTCTT 1907
Db 301 TGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGAGGATCCCGTGTCTCTT 360
QY 1908 TTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAGATACCTTT 1967
Db 361 TTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAGATACCTTT 420
QY 1968 TCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAAGCGCAGGCTGGGCTCTCTG 2027
Db 421 TCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAAGCGCAGGCTGGGCTCTCTG 480
QY 2028 ATGATTCTATCCAA 2041
Db 481 ATGATTCTATCCAA 494

RESULT 3
BG719710
LOCUS
DEFINITION
602690430F1 NIH_MGC_97 Homo sapiens cDNA clone IMAGE:4822807 5',
mRNA sequence.
BG719710
VERSION
BG719710.1 GI:13998897
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
1 (bases 1 to 770)
NIH-MGC http://mgc.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Miklos Palkovits, M.D., Ph.D.
CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki
Toshiyuki and Piero Carninci (RIKEN)
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10731 row: 1 column: 08
High quality sequence stop: 767.
FEATURES
source
1..770
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4822807"
/clone_lib="NIH_MGC_97"
/lab_host="DH10B"
/note="Organ: testis; Vector: pBluescriptR (modified
pBluescript KS+); Site_1: BamHI; Site_2: SalI-XhoI (gtcgag
); Oligo-dT primed using primer 5'-TTTTTTTTTTTNN-3',
size-selected for average insert size 2.2 kb and
normalized to 10^5. This is a primary library enriched
for full-length clones and constructed using the
Cap-trapper method (Carninci, in preparation). Library
constructed by M. Brownstein (NHGRI), National
Institutes of Health). Note: this is a NIH_MGC Library."
BASE COUNT 189 a 199 c 198 g 184 t
ORIGIN
Query Match 23.1%; Score 471.4; DB 12; Length 770;
Best Local Similarity 98.2%; Pred. No. 6.2e-104;
Matches 498; Conservative 0; Mismatches 6; Indels 3; Gaps 2;
2Y 1537 CCACGAGACTCAAAACAACTTGCTGGGACCATCCAAATGACAGAGCTCTACCACTTTT 1596
150 CCACGAGACTCAAAACAACTTGCTGGGACCATCCAAATGACAGAGCTCTACCACTTTT 209
2Y 1597 AGCTGACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAACTCCGAAGACT 1656
210 AGCTGACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAACTCCGAAGACT 269
2Y 1657 GCAGAAGGCCCTTTTGCTGGATCTCTTGAGCCTGTGACCTGCATGTCATGTCGCTGGACCA 1716
270 GCAGAAGGCCCTTTTGCTGGATCTCTTGAGCCTGTGACCTGCATGTCATGTCGCTGGACCA 329
2Y 1717 GCACAA--CCTCAAGCAAAATGACCAAGCCATGGATATCCTGCAGATTATTAATGTTG 1774
330 GCACAAATCCTCAAGTCAGATGACCAAGCCATGGATATCCTGCAGATTATTAATGTTG 389
2Y 1775 ACCACTATTATGACCGCTGGAGCAAGAGCACAAATTTGGTCAACGTCCTCTCTGC 1834
390 ACCACTATTATGACCGCTGGAGCAAGAGCACAAATTTGGTCAACGTCCTCTCTGC 449
2Y 1835 GTGGATATGTGTGTAAGTGGCTGCTGAATGTTTATGATACGGGACGAACAGGAGGATC 1894

Db 450 GTGGATATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGAGGATC 509
QY 1895 CGTGTCTCTCTTTTAAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAG 1954
Db 510 CGTGTCTCTCTTTTAAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAG 569
QY 1955 TACAGATACCTTTTCAAGCAAGTGGCAAGTTTCAACAGGATTTTGTGACCAGCGCAGGCTG 2014
Db 570 TACAGATACCTTTTCAAGCAAGTGGCAAGTTTCAACAGGATTTTGTGACCAGCGCAGGCTG 628
QY 2015 GGCTCTCTCTGCATGATTCTATCCAA 2041
Db 629 GGCTCTCTCTGCATGATTCTATCCAA 655
RESULT 4
BC009242
LOCUS
DEFINITION
BC009242 1490 bp mRNA linear HTC 12-JUL-2001
Homo sapiens, dystrophin (muscular dystrophy, Duchenne and Becker
types), includes DXS142, DXS164, DXS206, DXS230, DXS239, DXS268,
DXS269, DXS270, DXS272, clone IMAGE:3029414, mRNA.
ACCESSION
BC009242
VERSION
BC009242.1 GI:14714379
KEYWORDS
HTC.
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE
1 (bases 1 to 1490)
AUTHORS
Strausberg, R.
TITLE
Direct Submission
JOURNAL
Submitted (06-JUN-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
http://www.systemsbio.org
Contact: amadan@systemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia
Greene, Mark Kettelman and Anuradha Madan
Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 10 Row: j Column: 10
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 5032280
This clone has the following problem: incomplete processing.
FEATURES
Location/Qualifiers
source
1..1490
/organism="Homo sapiens"
/db_xref="LocusID:1756"
/db_xref="taxon:9606"
/clone="IMAGE:3029414"
/tissue_type="Muscle, rhabdomyosarcoma"
/clone_lib="NIH_MGC_17"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
BASE COUNT 505 a 299 c 328 g 358 t
ORIGIN
Query Match 22.3%; Score 454.6; DB 11; Length 1490;
Best Local Similarity 99.1%; Pred. No. 9e-100;
Matches 457; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 TCCTTACAGCAATTTGGAAGCTCTGAAGACAAGTCAATTTGGCAGTTTCATTTGATGGAGAG 60
|||||

URL:http://genome.gsc.riken.go.jp/ Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y. Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000) wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y. Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001) Please visit our web site (http://genome.gsc.riken.go.jp) for further details. e mouse tissues. Location/Qualifiers 1..704 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="1810074E05" /clone_lib="RIKEN full-length enriched, 10 day old male pancreas" /sex="male" /tissue_type="pancreas" /dev_stage="10 day old" /note="vector: pSPORT1; Site_1: Sali; Site_2: NotI; This clone is among a rearrayed set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos , and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the x chromosome, 1998, Hum Mol Genet 7: 1967-1978."									
BASE COUNT 173 a 185 c 167 g 179 t ORIGIN									
Query Match 22.0%; Score 449; DB 10; Length 704; Best Local Similarity 93.1%; Pred. No. 1.7e-98; Matches 470; Conservative 0; Mismatches 35; Indels 0; Gaps 0;									
QY	1537	CCACGAGACTCAAAACAACTTGCTGGGACCATCCAAAATGACAGAGCTCTACCAGTCTTT	1596						
Db	137	CCACGAGACCCAAACCACTTGTGGGACCAACCCAAAATGACAGAGCTCTACCAGTCTTT	196						
QY	1597	AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	1656						
Db	197	AGCTGACCTGAATAATGTTCAGGTTCTCCGCGTATAGGACTGCCATGAAGCTCAGAAGGCT	256						
QY	1657	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGGACCA	1716						
FEATURES source									
Location/Qualifiers 1..704 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="1810074E05" /clone_lib="RIKEN full-length enriched, 10 day old male pancreas" /sex="male" /tissue_type="pancreas" /dev_stage="10 day old" /note="vector: pSPORT1; Site_1: Sali; Site_2: NotI; This clone is among a rearrayed set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos , and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the x chromosome, 1998, Hum Mol Genet 7: 1967-1978."									
e mouse tissues. Further details. Please visit our web site (http://genome.gsc.riken.go.jp) for Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001) Computational Analysis of Full-Length Mouse cDNAs Compared with Hayashizaki,Y. Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Genes. Genome Res. 10 (10), 1617-1630 (2000) wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y. RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000) Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara Y. and Hayashizaki,Y. Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001) Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y. Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001) Please visit our web site (http://genome.gsc.riken.go.jp) for further details. e mouse tissues. Location/Qualifiers 1..704 /organism="Mus musculus" /strain="C57BL/6J" /db_xref="taxon:10090" /clone="1810074E05" /clone_lib="RIKEN full-length enriched, 10 day old male pancreas" /sex="male" /tissue_type="pancreas" /dev_stage="10 day old" /note="vector: pSPORT1; Site_1: Sali; Site_2: NotI; This clone is among a rearrayed set of 15,247 clones from 11 embryo cDNA libraries (including preimplantation stage embryos from unfertilized egg to blastocyst, embryonic part of E7.5 embryos, extraembryonic part of E7.5 embryos , and E12.5 female mesonephros/gonad) and one newborn ovary cDNA library. Average insert size 1.5 kb. All source libraries are cloned unidirectionally with Oligo(dT)-Not primers. References include: (1) Genome-wide expression profiling of mid-gestation placenta and embryo using a 15,000 mouse developmental cDNA microarray, 2000, Proc. Natl. Acad. Sci. U S A, 97: 9127-9132; (2) Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development, 2000, Development, 127: 1737-1749; (3) Genome-wide mapping of unselected transcripts from extraembryonic tissue of 7.5-day mouse embryos reveals enrichment in the t-complex and under-representation on the x chromosome, 1998, Hum Mol Genet 7: 1967-1978."									
BASE COUNT 173 a 185 c 167 g 179 t ORIGIN									
Query Match 22.0%; Score 449; DB 10; Length 704; Best Local Similarity 93.1%; Pred. No. 1.7e-98; Matches 470; Conservative 0; Mismatches 35; Indels 0; Gaps 0;									
QY	1537	CCACGAGACTCAAAACAACTTGCTGGGACCATCCAAAATGACAGAGCTCTACCAGTCTTT	1596						
Db	137	CCACGAGACCCAAACCACTTGTGGGACCAACCCAAAATGACAGAGCTCTACCAGTCTTT	196						
QY	1597	AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	1656						
Db	197	AGCTGACCTGAATAATGTTCAGGTTCTCCGCGTATAGGACTGCCATGAAGCTCAGAAGGCT	256						
QY	1657	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGGACCA	1716						
BASE COUNT 173 a 185 c 167 g 179 t ORIGIN									
Query Match 22.0%; Score 449; DB 10; Length 704; Best Local Similarity 93.1%; Pred. No. 1.7e-98; Matches 470; Conservative 0; Mismatches 35; Indels 0; Gaps 0;									
QY	1537	CCACGAGACTCAAAACAACTTGCTGGGACCATCCAAAATGACAGAGCTCTACCAGTCTTT	1596						
Db	137	CCACGAGACCCAAACCACTTGTGGGACCAACCCAAAATGACAGAGCTCTACCAGTCTTT	196						
QY	1597	AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	1656						
Db	197	AGCTGACCTGAATAATGTTCAGGTTCTCCGCGTATAGGACTGCCATGAAGCTCAGAAGGCT	256						
QY	1657	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGGACCA	1716						
BASE COUNT 173 a 185 c 167 g 179 t ORIGIN									
Query Match 22.0%; Score 449; DB 10; Length 704; Best Local Similarity 93.1%; Pred. No. 1.7e-98; Matches 470; Conservative 0; Mismatches 35; Indels 0; Gaps 0;									
QY	1537	CCACGAGACTCAAAACAACTTGCTGGGACCATCCAAAATGACAGAGCTCTACCAGTCTTT	1596						
Db	137	CCACGAGACCCAAACCACTTGTGGGACCAACCCAAAATGACAGAGCTCTACCAGTCTTT	196						
QY	1597	AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	1656						
Db	197	AGCTGACCTGAATAATGTTCAGGTTCTCCGCGTATAGGACTGCCATGAAGCTCAGAAGGCT	256						
QY	1657	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGGACCA	1716						
BASE COUNT 173 a 185 c 167 g 179 t ORIGIN									
Query Match 22.0%; Score 449; DB 10; Length 704; Best Local Similarity 93.1%; Pred. No. 1.7e-98; Matches 470; Conservative 0; Mismatches 35; Indels 0; Gaps 0;									
QY	1537	CCACGAGACTCAAAACAACTTGCTGGGACCATCCAAAATGACAGAGCTCTACCAGTCTTT	1596						
Db	137	CCACGAGACCCAAACCACTTGTGGGACCAACCCAAAATGACAGAGCTCTACCAGTCTTT	196						
QY	1597	AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	1656						
Db	197	AGCTGACCTGAATAATGTTCAGGTTCTCCGCGTATAGGACTGCCATGAAGCTCAGAAGGCT	256						
QY	1657	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGGACCA	1716						
BASE COUNT 173 a 185 c 167 g 179 t ORIGIN									
Query Match 22.0%; Score 449; DB 10; Length 704; Best Local Similarity 93.1%; Pred. No. 1.7e-98; Matches 470; Conservative 0; Mismatches 35; Indels 0; Gaps 0;									
QY	1537	CCACGAGACTCAAAACAACTTGCTGGGACCATCCAAAATGACAGAGCTCTACCAGTCTTT	1596						
Db	137	CCACGAGACCCAAACCACTTGTGGGACCAACCCAAAATGACAGAGCTCTACCAGTCTTT	196						
QY	1597	AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	1656						
Db	197	AGCTGACCTGAATAATGTTCAGGTTCTCCGCGTATAGGACTGCCATGAAGCTCAGAAGGCT	256						
QY	1657	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGGACCA	1716						
BASE COUNT 173 a 185 c 167 g 179 t ORIGIN									
Query Match 22.0%; Score 449; DB 10; Length 704; Best Local Similarity 93.1%; Pred. No. 1.7e-98; Matches 470; Conservative 0; Mismatches 35; Indels 0; Gaps 0;									
QY	1537	CCACGAGACTCAAAACAACTTGCTGGGACCATCCAAAATGACAGAGCTCTACCAGTCTTT	1596						
Db	137	CCACGAGACCCAAACCACTTGTGGGACCAACCCAAAATGACAGAGCTCTACCAGTCTTT	196						
QY	1597	AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	1656						
Db	197	AGCTGACCTGAATAATGTTCAGGTTCTCCGCGTATAGGACTGCCATGAAGCTCAGAAGGCT	256						
QY	1657	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGGACCA	1716						
BASE COUNT 173 a 185 c 167 g 179 t ORIGIN									
Query Match 22.0%; Score 449; DB 10; Length 704; Best Local Similarity 93.1%; Pred. No. 1.7e-98; Matches 470; Conservative 0; Mismatches 35; Indels 0; Gaps 0;									
QY	1537	CCACGAGACTCAAAACAACTTGCTGGGACCATCCAAAATGACAGAGCTCTACCAGTCTTT	1596						
Db	137	CCACGAGACCCAAACCACTTGTGGGACCAACCCAAAATGACAGAGCTCTACCAGTCTTT	196						
QY	1597	AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	1656						
Db	197	AGCTGACCTGAATAATGTTCAGGTTCTCCGCGTATAGGACTGCCATGAAGCTCAGAAGGCT	256						
QY	1657	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGGACCA	1716						
BASE COUNT 173 a 185 c 167 g 179 t ORIGIN									
Query Match 22.0%; Score 449; DB 10; Length 704; Best Local Similarity 93.1%; Pred. No. 1.7e-98; Matches 470; Conservative 0; Mismatches 35; Indels 0; Gaps 0;									
QY	1537	CCACGAGACTCAAAACAACTTGCTGGGACCATCCAAAATGACAGAGCTCTACCAGTCTTT	1596						
Db	137	CCACGAGACCCAAACCACTTGTGGGACCAACCCAAAATGACAGAGCTCTACCAGTCTTT	196						
QY	1597	AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	1656						
Db	197	AGCTGACCTGAATAATGTTCAGGTTCTCCGCGTATAGGACTGCCATGAAGCTCAGAAGGCT	256						
QY	1657	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGGACCA	1716						
BASE COUNT 173 a 185 c 167 g 179 t ORIGIN									
Query Match 22.0%; Score 449; DB 10; Length 704; Best Local Similarity 93.1%; Pred. No. 1.7e-98; Matches 470; Conservative 0; Mismatches 35; Indels 0; Gaps 0;									
QY	1537	CCACGAGACTCAAAACAACTTGCTGGGACCATCCAAAATGACAGAGCTCTACCAGTCTTT	1596						
Db	137	CCACGAGACCCAAACCACTTGTGGGACCAACCCAAAATGACAGAGCTCTACCAGTCTTT	196						
QY	1597	AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	1656						
Db	197	AGCTGACCTGAATAATGTTCAGGTTCTCCGCGTATAGGACTGCCATGAAGCTCAGAAGGCT	256						
QY	1657	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGGACCA	1716						
BASE COUNT 173 a 185 c 167 g 179 t ORIGIN									
Query Match 22.0%; Score 449; DB 10; Length 704; Best Local Similarity 93.1%; Pred. No. 1.7e-98; Matches 470; Conservative 0; Mismatches 35; Indels 0; Gaps 0;									
QY	1537	CCACGAGACTCAAAACAACTTGCTGGGACCATCCAAAATGACAGAGCTCTACCAGTCTTT	1596						
Db	137	CCACGAGACCCAAACCACTTGTGGGACCAACCCAAAATGACAGAGCTCTACCAGTCTTT	196						
QY	1597	AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	1656						
Db	197	AGCTGACCTGAATAATGTTCAGGTTCTCCGCGTATAGGACTGCCATGAAGCTCAGAAGGCT	256						
QY	1657	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGGACCA	1716						
BASE COUNT 173 a 185 c 167 g 179 t ORIGIN									
Query Match 22.0%; Score 449; DB 10; Length 704; Best Local Similarity 93.1%; Pred. No. 1.7e-98; Matches 470; Conservative 0; Mismatches 35; Indels 0; Gaps 0;									
QY	1537	CCACGAGACTCAAAACAACTTGCTGGGACCATCCAAAATGACAGAGCTCTACCAGTCTTT	1596						
Db	137	CCACGAGACCCAAACCACTTGTGGGACCAACCCAAAATGACAGAGCTCTACCAGTCTTT	196						
QY	1597	AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	1656						
Db	197	AGCTGACCTGAATAATGTTCAGGTTCTCCGCGTATAGGACTGCCATGAAGCTCAGAAGGCT	256						
QY	1657	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGGACCA	1716						
BASE COUNT 173 a 185 c 167 g 179 t ORIGIN									
Query Match 22.0%; Score 449; DB 10; Length 704; Best Local Similarity 93.1%; Pred. No. 1.7e-98; Matches 470; Conservative 0; Mismatches 35; Indels 0; Gaps 0;									
QY	1537	CCACGAGACTCAAAACAACTTGCTGGGACCATCCAAAATGACAGAGCTCTACCAGTCTTT	1596						
Db	137	CCACGAGACCCAAACCACTTGTGGGACCAACCCAAAATGACAGAGCTCTACCAGTCTTT	196						
QY	1597	AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	1656						
Db	197	AGCTGACCTGAATAATGTTCAGGTTCTCCGCGTATAGGACTGCCATGAAGCTCAGAAGGCT	256						
QY	1657	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGGACCA	1716						
BASE COUNT 173 a 185 c 167 g 179 t ORIGIN									
Query Match 22.0%; Score 449; DB 10; Length 704; Best Local Similarity 93.1%; Pred. No. 1.7e-98; Matches 470; Conservative 0; Mismatches 35; Indels 0; Gaps 0;									
QY	1537	CCACGAGACTCAAAACAACTTGCTGGGACCATCCAAAATGACAGAGCTCTACCAGTCTTT	1596						
Db	137	CCACGAGACCCAAACCACTTGTGGGACCAACCCAAAATGACAGAGCTCTACCAGTCTTT	196						
QY	1597	AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	1656						
Db	197	AGCTGACCTGAATAATGTTCAGGTTCTCCGCGTATAGGACTGCCATGAAGCTCAGAAGGCT	256						
QY	1657	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGGACCA	1716						
BASE COUNT 173 a 185 c 167 g 179 t ORIGIN									
Query Match 22.0%; Score 449; DB 10; Length 704; Best Local Similarity 93.1%; Pred. No. 1.7e-98; Matches 470; Conservative 0; Mismatches 35; Indels 0; Gaps 0;									
QY	1537	CCACGAGACTCAAAACAACTTGCTGGGACCATCCAAAATGACAGAGCTCTACCAGTCTTT	1596						
Db	137	CCACGAGACCCAAACCACTTGTGGGACCAACCCAAAATGACAGAGCTCTACCAGTCTTT	196						
QY	1597	AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	1656						
Db	197	AGCTGACCTGAATAATGTTCAGGTTCTCCGCGTATAGGACTGCCATGAAGCTCAGAAGGCT	256						
QY	1657	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGGACCA	1716						
BASE COUNT 173 a 185 c 167 g 179 t ORIGIN									
Query Match 22.0%; Score 449; DB 10; Length 704; Best Local Similarity 93.1%; Pred. No. 1.7e-98; Matches 470; Conservative 0; Mismatches 35; Indels 0; Gaps 0;									
QY	1537	CCACGAGACTCAAAACAACTTGCTGGGACCATCCAAAATGACAGAGCTCTACCAGTCTTT	1596						
Db	137	CCACGAGACCCAAACCACTTGTGGGACCAACCCAAAATGACAGAGCTCTACCAGTCTTT	196						
QY	1597	AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	1656						
Db	197	AGCTGACCTGAATAATGTTCAGGTTCTCCGCGTATAGGACTGCCATGAAGCTCAGAAGGCT	256						
QY	1657	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGGACCA	1716						
BASE COUNT 173 a 185 c 167 g 179 t ORIGIN									
Query Match 22.0%; Score 449; DB 10; Length 704; Best Local Similarity 93.1%; Pred. No. 1.7e-98; Matches 470; Conservative 0; Mismatches 35; Indels 0; Gaps 0;									
QY	1537	CCACGAGACTCAAAACAACTTGCTGGGACCATCCAAAATGACAGAGCTCTACCAGTCTTT	1596						
Db	137	CCACGAGACCCAAACCACTTGTGGGACCAACCCAAAATGACAGAGCTCTACCAGTCTTT	196						
QY	1597	AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	1656						
Db	197	AGCTGACCTGAATAATGTTCAGGTTCTCCGCGTATAGGACTGCCATGAAGCTCAGAAGGCT	256						
QY	1657	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGGACCA	1716						
BASE COUNT 173 a 185 c 167 g 179 t ORIGIN									
Query Match 22.0%; Score 449; DB 10; Length 704; Best Local Similarity 93.1%; Pred. No. 1.7e-98; Matches 470; Conservative 0; Mismatches 35; Indels 0; Gaps 0;									
QY	1537	CCACGAGACTCAAAACAACTTGCTGGGACCATCCAAAATGACAGAGCTCTACCAGTCTTT	1596						
Db	137	CCACGAGACCCAAACCACTTGTGGGACCAACCCAAAATGACAGAGCTCTACCAGTCTTT	196						
QY	1597	AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	1656						
Db	197	AGCTGACCTGAATAATGTTCAGGTTCTCCGCGTATAGGACTGCCATGAAGCTCAGAAGGCT	256						
QY	1657	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGGACCA	1716						
BASE COUNT 173 a 185 c 167 g 179 t ORIGIN									
Query Match 22.0%; Score 449; DB 10; Length 704; Best Local Similarity 93.1%; Pred. No. 1.7e-98; Matches 470; Conservative 0; Mismatches 35; Indels 0; Gaps 0;									
QY	1537	CCACGAGACTCAAAACAACTTGCTGGGACCATCCAAAATGACAGAGCTCTACCAGTCTTT	1596						
Db	137	CCACGAGACCCAAACCACTTGTGGGACCAACCCAAAATGACAGAGCTCTACCAGTCTTT	196						
QY	1597	AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	1656						
Db	197	AGCTGACCTGAATAATGTTCAGGTTCTCCGCGTATAGGACTGCCATGAAGCTCAGAAGGCT	256						
QY	1657	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGGACCA	1716						
BASE COUNT 173 a 185 c 167 g 179 t ORIGIN									
Query Match 22.0%; Score 449; DB 10; Length 704; Best Local Similarity 93.1%; Pred. No. 1.7e-98; Matches 470; Conservative 0; Mismatches 35; Indels 0; Gaps 0;									
QY	1537	CCACGAGACTCAAAACAACTTGCTGGGACCATCCAAAATGACAGAGCTCTACCAGTCTTT	1596						
Db	137	CCACGAGACCCAAACCACTTGTGGGACCAACCCAAAATGACAGAGCTCTACCAGTCTTT	196						
QY	1597	AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT	1656						
Db	197	AGCTGACCTGAATAATGTTCAGGTTCTCCGCGTATAGGACTGCCATGAAGCTCAGAAGGCT	256						
QY	1657	GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGGACCA	1716						
BASE COUNT 173 a 185 c 167 g 179 t ORIGIN									
Query Match 22.0%; Score 449; DB 10; Length 704;									


```
RESULT 14
AL672616          608 bp      mRNA      linear      EST 18-MAR-2002
LOCUS
DEFINITION
AL672616 XGC-gastrula Silurana tropicalis cDNA clone TGas054h08 5',
mRNA sequence.
ACCESSION
VERSION          AL672616.1  GI:19528972
KEYWORDS
SOURCE
ORGANISM
western clawed frog.
Silurana tropicalis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Silurana.
1 (bases 1 to 608)
Taylor,R., Ashurst,J.L., Croning,M.D.R., Zorn,A.M. and Rogers,J.
Sanger Xenopus tropicalis EST project 2002
Unpublished (2001)
JOURNAL
COMMENT
Contact: Taylor R
Sanger Centre
Hinxtan, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TGas054h08.plcSP6
Sequencing primer: PLCSP6
This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.
Location/Qualifiers
1. 608
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="TGas054h08"
/clone_lib="XGC-gastrula"
/dev_stage="gastrula (stages 10.5-13 mixed)"
/lab_host="Escherichia coli XL1-blue"
/note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA
was oligo dT primed from 5ug of poly A+ RNA from stages
10-13 gastrulae. EcoRI-NotI cut cDNA was then ligated
into pCS107 with EcoRI at the 5' end and NotI at the 3'
end."
BASE COUNT      150 a      155 c      151 g      152 t
ORIGIN
Query Match      16.2%; Score 330; DB 9; Length 608;
Best Local Similarity 85.0%; Pred. No. 1.6e-69;
Matches 369; Conservative 0; Mismatches 65; Indels 0; Gaps 0;
QY 1537 CCACGAGACTCAAACAACCTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCACTCTT 1596
||| ||||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| |||||
Db 175 CCATGAGACACAAACCCTGCTGGGATCATCCCAAAATGACAGAAATTATACCAATCTT 234
QY 1597 AGCTGACCTGAATAATGTGAGATTCTCAGCTTATAGGACTGCCATGAACTCCGAAGACT 1656
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 235 AGCTGACCTGAACAATGTGCGATTCTCAGCATACAGAACTGCCATGAAGCTAAGGAGATT 294
QY 1657 GCAGAAGGCCCTTTGCTTGGATCTTTGAGCCTGTGAGCTGCATGTGATGCTTGGACCA 1716
||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| ||| ||||| |||
Db 295 GCAAAAGGCCCTTGCTGCTGGATTGCTAGGGCTGTCTGAGCTTGTGAAGCCTTGGACCA 354
QY 1717 GCACAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTGAC 1776
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 355 GCACAACCTGAAGCAGAAATGACCAGCTGATGGACATCCTGCAGATTATTAATTGCTTGAC 414
QY 1777 CACTATTATGACCGCTGGAGCAAGAGCACACAATTTGGTCAACGTCCTCTCTCGCT 1836
||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 415 CACAATTATGATCGACTGGAGCAAGAGCACAAATAATCTGGTGAACGTTCTCTCTCGCT 474
QY 1837 GGATATGTGCTGACTGGCTGCTGAATGTTTATGATACGGGACGAACAGGAGGATCCG 1896
||| ||||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 475 GGACATGTGCCTCAACTGGCTGCTGAATGTTTATGACACGGGTGGAACGGGACGTATACG 534
QY 1897 TGTCTGTCTTTTAAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAAGTA 1956
```

```
||| | ||||| ||||| | ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 535 CGTCTTATCTTTTAAAACTGGTGTAAATTTCCCTGTGTAAGCACATTTGGAAGATAAGTA 594
QY 1957 CAGATACCTTTTCA 1970
||||| | ||||
Db 595 CAGATACCTTATTCA 608
RESULT 15
BE334408
LOCUS
DEFINITION
BE334408 532 bp      mRNA      linear      EST 14-JUL-2000
ug97g06.y1 Soares_mammary_gland_NMLMG Mus musculus cDNA clone
IMAGE:1550362 5' similar to gb:M68859 Mouse dystrophin mRNA, exons
1-7 and complete cds (MOUSE);, mRNA sequence.
ACCESSION
VERSION          BE334408.1  GI:9208184
KEYWORDS
SOURCE
ORGANISM
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 532)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
JOURNAL
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
This clone is available royalty-free through LLNL; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:955710
Seq primer: -40RP from Gibco
High quality sequence stop: 449.
Location/Qualifiers
1. 532
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="IMAGE:1550362"
/clone_lib="Soares_mammary_gland_NMLMG"
/sex="female (lactating)"
/tissue_type="mammary gland"
/lab_host="DH10B"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; 1st strand cDNA was prepared from mammary
gland tissue from a lactating female, and was then primed
with a Not I - oligo(dT) primer. Double-stranded cDNA was
ligated to Eco RI adaptors (Pharmacia), digested with Not
I and cloned into the Not I and Eco RI sites of the
modified pT7T3 vector. Library is normalized. Library
was constructed by Bento Soares and M. Fatima Bonaldo."
BASE COUNT      121 a      151 c      132 g      128 t
ORIGIN
Query Match      16.1%; Score 329; DB 10; Length 532;
Best Local Similarity 93.2%; Pred. No. 2.6e-69;
Matches 344; Conservative 0; Mismatches 25; Indels 0; Gaps 0;
QY 1537 CCACGAGACTCAAACAACCTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCACTCTT 1596
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 164 CCACGAGACCCCAACCCTTGTTGGGACCAACCCCAAAATGACAGAGCTCTACCACTCTT 223
QY 1597 AGCTGACCTGAATAATGTGAGATTCTCAGCTTATAGGACTGCCATGAACTCCGAAGACT 1656
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 224 AGCTGACCTGAATAATGTGAGTTCTCCGCTATAGGACTGCCATGAAGCTCAGAAGGCT 283
QY 1657 GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTGAGCTGCATGCTGATGCCCTGGACCA 1716
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 284 CCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTGAGCTGCATGCTGATGCCCTGGACCA 343
QY 1717 GCACAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTGAC 1776
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 344 GCACAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATAATTAACCTGTTGAC 403
```


★

Job time : 1950.12 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 08:10:12 ; Search time 46.4852 Seconds
(without alignments)
13465.096 Million cell updates/sec

Title: US-09-845-416-12_COPY_960_3000
Perfect score: 2041
Sequence: 1 tccttcacagcatttgaag.....ttctgcatgattctatccaa 2041

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 875442

Minimum DB seq length: 0
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA:*
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	40.4	2.0	2223	1	US-08-257-073-4
2	38.6	1.9	1995	1	US-08-425-069-3
3	38.6	1.9	1995	2	US-08-317-844B-3
4	38.2	1.9	1131	6	5180810-3
5	38.2	1.9	1784	6	5180810-2
6	38	1.9	1394	4	US-09-247-155-76
7	36.8	1.8	1886	6	5210183-1
8	36.6	1.8	1845	4	US-08-887-534A-22
9	36.2	1.8	289	4	US-09-007-005-17
10	36.2	1.8	289	4	US-09-244-796-17
11	36.2	1.8	2447	2	US-09-014-969-14
12	36	1.8	2873	4	US-08-630-915A-193
13	35.8	1.8	790	4	US-09-461-474-13
14	35.6	1.7	4868	1	US-08-139-937-12
15	35.6	1.7	4868	5	PCT-US93-11310-12
16	35.2	1.7	3466	1	US-08-468-036-38
17	35.2	1.7	3466	2	US-08-376-843-38
18	35	1.7	4766	4	PCT-US93-07261-10
19	34.8	1.7	2442	4	US-09-040-485-1
20	34.8	1.7	3337	1	US-08-072-610-1
21	34.8	1.7	3337	2	US-08-719-822B-1
22	34.8	1.7	3337	2	US-09-092-458-1
23	34.6	1.7	198	5	PCT-US95-10668-3
24	34.6	1.7	198	5	PCT-US95-10668-4
25	34.6	1.7	2277	1	US-08-676-967-5
26	34.6	1.7	2277	1	US-08-676-974-5
27	34.6	1.7	2277	2	US-09-098-487-5

C	28	34.6	1.7	2704	4	US-08-857-076-44	Sequence 44, Appl
C	29	34.6	1.7	3499	4	US-08-857-076-43	Sequence 43, Appl
	30	34.4	1.7	277	4	US-09-007-005-3	Sequence 3, Appli
	31	34.4	1.7	277	4	US-09-244-796-3	Sequence 3, Appli
	32	34.4	1.7	717	4	US-09-461-697-189	Sequence 189, App
	33	34.4	1.7	774	4	US-09-461-697-187	Sequence 187, App
	34	34.4	1.7	819	4	US-09-461-697-185	Sequence 185, App
	35	34.4	1.7	1047	4	US-09-671-950-1	Sequence 1, Appli
	36	34.4	1.7	1047	4	US-09-671-950-3	Sequence 3, Appli
	37	34.4	1.7	1047	4	US-09-671-950-5	Sequence 5, Appli
	38	34.4	1.7	1047	4	US-09-671-950-7	Sequence 7, Appli
	39	34.4	1.7	1047	4	US-09-671-950-9	Sequence 9, Appli
	40	34.4	1.7	1047	4	US-09-671-950-11	Sequence 11, Appl
	41	34.4	1.7	1047	4	US-09-671-950-13	Sequence 13, Appl
	42	34.4	1.7	1669	4	US-09-461-697-184	Sequence 184, App
C	43	34.2	1.7	621	4	US-09-328-111-204	Sequence 204, App
C	44	34.2	1.7	734	3	US-09-014-583-1	Sequence 1, Appli
	45	34.2	1.7	750	4	US-08-961-527-370	Sequence 370, App

ALIGNMENTS

RESULT 1
US-08-257-073-4
; Sequence 4, Application US/08257073
; Patent No. 5766597
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: de Taisne, Charles
; APPLICANT: Tine, John A.
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: New York
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/257,073
; FILING DATE: 09-JUN-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/075,783
; FILING DATE: 11-JUN-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/852,305
; FILING DATE: 18-MAR-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/672,183
; FILING DATE: 20-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2570
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066 CURTMS
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2223 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-257-073-4

100

LOCATION: 46..675
FEATURE:
NAME/KEY: sig_peptide
LOCATION: 46..87
OTHER INFORMATION: Von Heijne matrix
OTHER INFORMATION: score 5.9
OTHER INFORMATION: seq LTLGLSLILAGL/IV
FEATURE:
NAME/KEY: polyA_signal
LOCATION: 1363..1368
FEATURE:
NAME/KEY: polyA_site
LOCATION: 1382..1394
US-09-247-155-76

Query Match 1.9%; Score 38; DB 4; Length 1394;
Best Local Similarity 59.1%; Pred. No. 0.17;
Matches 65; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

QY 369 AAAACAAAGCAATTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACTGAAAGAGT 428
Db 1087 AAAACAAATTTAAATATTTCAAATTTTAAATTTGCTGCACCATAAGATGAATAAGAGC 1028
QY 429 TGAATGACTGGCTAACAAAAACAGAAAGAAACAAAGGAAAAATGGAGAA 478
Db 1027 TTAATAAGGAAAGAAAAACAAACAAACAAACAAACAAACAAACGATGGA 978

RESULT 7
5210183-1
Patent No. 5210183
APPLICANT: LINDAHL, GUNNAR;FRITHZ, ELISABET;HEDEN, LARS-OLOF
TITLE OF INVENTION: PROTEIN ARP, WITH IMMUNOGLOBULIN A
BINDING ACTIVITY, THE CORRESPONDING VECTORS AND HOSTS, REAGENT
KIT AND PHARMACEUTICAL COMPOSITION
NUMBER OF SEQUENCES: 3
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/270,099
FILING DATE: 14-NOV-1988
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 186,097
FILING DATE: 25-APR-1988
SEQ ID NO:1:
LENGTH: 1886

Query Match 1.8%; Score 36.8; DB 6; Length 1886;
Best Local Similarity 45.5%; Pred. No. 0.5;
Matches 131; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 274 AACAGGAAATATATCAGAAAGATGAAGAACTGAAGTACAAAGAGCAGATGAATCTCCTAAA 333
Db 261 AGCGTGGAAGTGGCCTAAAGAAATATAACCGCTTACTTAAGGAAATGAGGAGCTCAAGGT 320
QY 334 TTCAAGATGGGAATGCCTCAGGGTAGCTAGCATGGGAAACAAAGCAATTTACATAGAGT 393
Db 321 AGAAGTGAAGAAATATCTATCTTATGCTGACGATAAGAAAGATCCTCAATCTAGAGC 380
QY 394 TTTAATGGATCTCCAGAAATCAGAAACTGAAGAGTTGAATGACTGGCTAACAAAAACAGA 453
Db 381 ATTAATGGTGAAATCAAGATCTTCGAAAAAGAGAGGGGACAATATCAGGACAAATAGA 440
QY 454 AGAAGAAACAAAGGAAATGGAGGAGAGCCTCTTGGACCTGATCTTGAAGACCTAAACG 513
Db 441 AGAACTTGAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAGAAAG 500
QY 514 CCAAGTACAAACATAGAGTGTCTTCAAGAAAGATCTAGAACAAAGAA 561
Db 501 TCAATAGAGCAGATAGAGCATTATCAAGAAACAAACAAACAAACAAACATCA 548

RESULT 8
US-08-887-534A-22

Sequence 22, Application US/08887534A
Patent No. 6455323
GENERAL INFORMATION:
APPLICANT: Holden, David W.
TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS
NUMBER OF SEQUENCES: 106
CORRESPONDENCE ADDRESS:
ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
STREET: 233 South Wacker Drive/6300 Sears Tower
CITY: Chicago
STATE: Illinois
COUNTRY: United States of America
ZIP: 60606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/887,534A
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: Rin-Laures, Li-Hsien
REGISTRATION NUMBER: 33,547
REFERENCE/DOCKET NUMBER: 28341/33996
TELECOMMUNICATION INFORMATION:
TELEPHONE: (312) 474-6300
TELEFAX: (312) 474-0448
TELEX: (312) 474-6600
INFORMATION FOR SEQ ID NO: 22:
SEQUENCE CHARACTERISTICS:
LENGTH: 1845 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
DESCRIPTION: /desc = "DNA (genomic) (p14c15)"
FEATURE:
NAME/KEY: CDS
LOCATION: 3..788
FEATURE:
NAME/KEY: CDS
LOCATION: 856..1842
US-08-887-534A-22

Query Match 1.8%; Score 36.6; DB 4; Length 1845;
Best Local Similarity 48.0%; Pred. No. 0.57;
Matches 135; Conservative 0; Mismatches 144; Indels 2; Gaps 1;

QY 210 TGGATTTCAGAGCCCCATCAGGGCCGGTGGTAAATATCTACAATTTGGGAAGTAAGCTGA 269
Db 349 TGGCTTTGGATAAACATATGATGCATGTAACAATTAGTTATCCCTACCTGACAATCAGC 408
QY 270 TTGGAACAGGAAATATCAGAAAGATGAAGAACTGAAGTACAAAGAGCAGATGAATCTCC 329
Db 409 TACTTACCAACTATTTACGGAACCTTGAAGAGGTGCTGTAAATGTTGATATGATTTCAC 468
QY 330 TAAATTCAGATGGGAATGCCTCAGGGTAGCTAGCATGGGAAACAAAGCAATTTACATA 389
Db 469 AAATCGTCAACTTGGATGGGCTACAACATATCCTTCACGATTAAGATAGTATTTTCATC 528
QY 390 GAGTTTAAATGGATCTCCAGAAATCAGAACTGAAAGAGTTGAATGACTGGCTAACAAAA 449
Db 529 AAATTTCTATGATTTCTT--GAAACATTAAGAAATCAATATGAAGCATTAGCTTATAAAT 586
QY 450 CAGAAGAAAGAAAGAAATGGAGGAAAGAGCTCTTGA 490
Db 587 CAATGAGCATTATGTCAAAATTTTCATTAAATTTGGCTCAGGCA 627

RESULT 9
US-09-007-005-17


```

; Sequence 17, Application US/09007005B
; Patent No. 6258558
;
; GENERAL INFORMATION:
;
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
;
; TITLE OF INVENTION: SELECTION OF PROTEINS
; TITLE OF INVENTION: FUSIONS
;
; FILE REFERENCE: 00786/350003
;
; CURRENT APPLICATION NUMBER: US/09/007,005B
;
; CURRENT FILING DATE: 1998-01-14
;
; EARLIER APPLICATION NUMBER: 60/035,963
;
; EARLIER FILING DATE: 1997-01-27
;
; EARLIER APPLICATION NUMBER: 60/064,491
;
; EARLIER FILING DATE: 1997-11-06
;
; NUMBER OF SEQ ID NOS: 33
;
; SOFTWARE: FastSEQ for Windows Version 4.0
;
; SEQ ID NO 17

```

```

;
; LENGTH: 289
;
; TYPE: RNA
;
; ORGANISM: Artificial Sequence
;
; FEATURE:
;
; OTHER INFORMATION: Translation template
;
; FEATURE:
;
; NAME/KEY: misc_feature
;
; LOCATION: (1)..(289)
;
; OTHER INFORMATION: n = A,T,C or G
US-09-007-005-17

```

Query Match 1.8%; Score 36.2; DB 4; Length 289;
Best Local Similarity 6.1%; Pred. No. 0.22;
Matches 14; Conservative 99; Mismatches 118; Indels

Matches 14; Conservative 99; Mismatches 118; Indels 0; Gaps 0;

Qy	342	GGGAATGCCTCAGGTAAGTAGCATGGAAAAACAAAGCAATTTACATAGAGTTTTAATGG	401
Dd	3	RGRGRACRARURURARCRURARURURARCRARURURARCRARURURGRNRNRS	62
Qy	402	ATCTCCAGAATCAGAAACTGAAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGAA	461
Dd	63	RNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS	122
Qy	462	CAAGGAAAATGGAGGAAGCCCTCTTGGACCTGATCTTGAAGACCTAAACGCCAAGTAC	521
Dd	123	RNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS	182
Qy	522	AACAACATAAGTGCTTCAGAAGATCTAGAACAAAGACAGGGGTCA	572
Dd	183	RNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRSRNRNRS	233

RESULT 10

```

US-09-244-796-17
; Sequence 17, Application US/092444796
; Patent No. 6281344
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; TITLE OF INVENTION: FUSIONS
; FILE REFERENCE: 00786/350007
; CURRENT APPLICATION NUMBER: US/09/244,796
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; EARLIER APPLICATION NUMBER: 09/007,005
; EARLIER FILING DATE: 1998-01-14
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289

```

```

; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(289)
; OTHER INFORMATION: n = A,T,C or G
US-09-244-796-17

```

Query Match 1.8%; Score 36.2; DB 4; Length 289;
Best Local Similarity 6.1%; Pred. No. 0.22;
Matches 14; Conservative 99; Mismatches 118; Indels

Matches 14; Conservative 99; Mismatches 118; Indels 0; Gaps 0;

[illegible]

RESULT 11

```

US-09-014-969-14
; Sequence 14, Application US/09014969
; Patent No. 5965397
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Racie, Lisa A.
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; APPLICANT: Agostino, Michael J.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
; TITLE OF INVENTION: ENCODING THEM
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 CambridgePark Drive
; CITY: Cambridge
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/014,969
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Sprunger, Suzanne A.
; REGISTRATION NUMBER: 41,323
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 498-8284
; TELEFAX: (617) 876-5851
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2447 base pairs
; TYPE: nucleic acid

```



```

Query Match          1.8%; Score 35.8; DB 4; Length 790;
Best Local Similarity 61.1%; Pred. No. 0.57;
Matches 58; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 376 AAGCAATTTACATAGAGTTTTAATGGATCTCCAGAACTCAGAAAAGTGAATGA 435
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 690 ATGCAATTTTCAGACAGTTTTGTGAGCCTTGTGAGTTTATCAACCTGCACGTGTAGTTTCA 749
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 436 CTGGCTAACAAAAACAGAAAGAAACAAGGAAAA 470
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 750 GCGGCACAAAAAAGAAAAAAGAAAAAAGAAAAA 784
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
US-08-139-937-12
; Sequence 12, Application US/08139937
; Patent No. 5821070
; GENERAL INFORMATION:
; APPLICANT: LEE, WEN-HWA
; APPLICANT: SHAN, BEI
; TITLE OF INVENTION: CELLULAR GENES ENCODING
; TITLE OF INVENTION: RETINOBLASTOMA-ASSOCIATED PROTEINS
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CAMPBELL AND FLORES
; STREET: 4370 LA JOLLA VILLAGE DRIVE
; CITY: SAN DIEGO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS

```


GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OMnucleic - nucleic search, using sw model

Run on: April 25, 2003, 17:39:49 ; Search time 135.61 Seconds
(without alignments)
16376.941 Million cell updates/sec

Title: US-09-845-416-12_COPY_960_3000
Perfect score: 2041
Sequence: 1 tccttcacagcatttggaag.....ttctgcatgattctatccaa 2041

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 709820 seqs, 544064369 residues 1410746

Total number of hits satisfying chosen parameters: 1410746

Minimum DB seq length: 0
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	190.2	9.3	256	10	US-09-864-761-21956
2	177	8.7	466	10	US-09-864-761-6092
3	153.4	7.5	467	10	US-09-864-761-11083
4	151	7.4	151	10	US-09-864-761-27715
5	43	2.1	436	10	US-09-960-352-10742
6	42.6	2.1	449	9	US-09-918-995-24084
7	42.6	2.1	2247	10	US-09-960-253-157
8	42.2	2.1	423	10	US-09-864-761-18355
9	41.8	2.0	440	9	US-10-184-644-442
10	41.8	2.0	440	9	US-10-184-634-442
11	40.6	2.0	425	10	US-09-960-352-4010
12	40.6	2.0	747	9	US-10-123-155-212
13	40.4	2.0	428	10	US-09-864-864-204
14	39.8	2.0	418	10	US-09-960-352-12395
15	39.6	1.9	1381	9	US-09-822-846-343
16	39.6	1.9	1594	9	US-10-174-590-183
17	39.6	1.9	1594	9	US-10-176-758-183
18	39.6	1.9	1594	9	US-10-175-737-183
19	39.6	1.9	1594	9	US-10-173-706-183

C 20	39.6	1.9	1594	9	US-10-175-738-183	Sequence 183, App
C 21	39.6	1.9	1594	9	US-10-175-752-183	Sequence 183, App
C 22	39.6	1.9	1594	9	US-10-176-482-183	Sequence 183, App
C 23	39.6	1.9	1594	9	US-10-176-757-183	Sequence 183, App
C 24	39.6	1.9	1594	9	US-10-176-913-183	Sequence 183, App
C 25	39.6	1.9	1594	9	US-10-180-552-183	Sequence 183, App
C 26	39.6	1.9	1594	9	US-10-180-557-183	Sequence 183, App
C 27	39.6	1.9	1594	9	US-10-173-700-183	Sequence 183, App
C 28	39.6	1.9	1594	9	US-10-174-572-183	Sequence 183, App
C 29	39.6	1.9	1594	9	US-10-174-579-183	Sequence 183, App
C 30	39.6	1.9	1594	9	US-10-174-582-183	Sequence 183, App
C 31	39.6	1.9	1594	9	US-10-174-588-183	Sequence 183, App
C 32	39.6	1.9	1594	9	US-10-175-739-183	Sequence 183, App
C 33	39.6	1.9	1594	9	US-10-175-740-183	Sequence 183, App
C 34	39.6	1.9	1594	9	US-10-175-743-183	Sequence 183, App
C 35	39.6	1.9	1594	9	US-10-176-488-183	Sequence 183, App
C 36	39.6	1.9	1594	9	US-10-176-492-183	Sequence 183, App
C 37	39.6	1.9	1594	9	US-10-176-747-183	Sequence 183, App
C 38	39.6	1.9	1594	9	US-10-176-750-183	Sequence 183, App
C 39	39.6	1.9	1594	9	US-10-176-985-183	Sequence 183, App
C 40	39.6	1.9	1594	9	US-10-176-987-183	Sequence 183, App
C 41	39.6	1.9	1594	9	US-10-176-991-183	Sequence 183, App
C 42	39.6	1.9	1594	9	US-10-176-992-183	Sequence 183, App
C 43	39.6	1.9	1594	9	US-10-176-993-183	Sequence 183, App
C 44	39.6	1.9	1594	9	US-10-184-658-183	Sequence 183, App
C 45	39.6	1.9	1594	9	US-10-173-695-183	Sequence 183, App

ALIGNMENTS

RESULT 1
US-09-864-761-21956
; Sequence 21956, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecmica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21956
; LENGTH: 256
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004468.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
; OTHER INFORMATION: SWISSPROT HIT: P11532, EVALUE 2.00e-30
; OTHER INFORMATION: NT HIT: gi5032282, EVALUE 1.00e-103
; US-09-864-761-21956
Query Match 9.3%; Score 190.2; DB 10; Length 256;
Best Local Similarity 98.5%; Pred. No. 1.1e-47;
Matches 192; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
QY 11 CATTGGGAAGCTCCTGAAGACAAAGTCATTTGGCAGTTCATTGATGGAGAGTGAAGTAAAC 70
|||||
Db 1 CATTGGGAAGCTCCTGAAGACAAAGTCATTTGGCAGTTCATTGATGGAGAGTGAAGTAAAC 60
QY 71 CTGGACCGTTATCAACACAGCTTTAGAAAGATTTATCGTGGCTTCTTTCTGCTGAGGAC 130
|||||
Db 61 CTGGACCGTTATCAACACAGCTTTAGAAAGATTTATCGTGGCTTCTTTCTGCTGAGGAC 120
QY 131 ACATTGCAAGCACAAAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGACCAAGTTTCAT 190
|||||
Db 121 ACATTGCAAGCACAAAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGACCAAGTTTCAT 180
QY 191 ACTCATGAGGGGTAC 205
|||||
Db 181 ACTCATGAGGGTAAAC 195
RESULT 2
US-09-864-761-6092
; Sequence 6092, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27

; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 6092
; LENGTH: 466
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004468.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
; US-09-864-761-6092
Query Match 8.7%; Score 177; DB 10; Length 466;
Best Local Similarity 100.0%; Pred. No. 1.8e-43;
Matches 177; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 8 CAGCATTTGGGAAGCTCCTGAAGACAAAGTCATTTGGCAGTTCATTGATGGAGAGTGAAGTA 67
|||||
Db 290 CAGCATTTGGGAAGCTCCTGAAGACAAAGTCATTTGGCAGTTCATTGATGGAGAGTGAAGTA 349
QY 68 AACCTGGACCGTTATCAACACAGCTTTAGAAAGATTTATCGTGGCTTCTTTCTGCTGAG 127
|||||
Db 350 AACCTGGACCGTTATCAACACAGCTTTAGAAAGATTTATCGTGGCTTCTTTCTGCTGAG 409
QY 128 GACACATTGCAAGCACAAAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGACCAAG 184
|||||
Db 410 GACACATTGCAAGCACAAAGGAGAGATTTCTAATGATGTGGAAGTGGTGAAGACCAAG 466
RESULT 3
US-09-864-761-11083
; Sequence 11083, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aeomica-x-1

CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 11083
LENGTH: 467
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC004468.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.82
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.89
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.84
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.96
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.88
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.81
JS-09-864-761-11083

Query Match 7.5%; Score 153.4; DB 10; Length 467;
Best Local Similarity 99.4%; Pred. No. 2.9e-36;
Matches 154; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 379 CAATTACATAGAGTTTAAATGGATCTCCAGAACTCAGAAACTGAAAGAGTTGAATGACTG 438
|||
Db 313 CAGTTACATAGAGTTTAAATGGATCTCCAGAACTCAGAAACTGAAAGAGTTGAATGACTG 372
QY 439 GCTAACAAAACAGAGAAGAAACAAGGAAATGGAGGAGCCCTCTTGGACCTGATCT 498
|||
Db 373 GCTAACAAAACAGAGAAGAAACAAGGAAATGGAGGAGCCCTCTTGGACCTGATCT 432

QY 499 TGAAGACCTAAACCGCAAGTACAACAATAAGG 533
|||
Db 433 TGAAGACCTAAACCGCAAGTACAACAATAAGG 467

RESULT 4
JS-09-864-761-27715

; Sequence 27715, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 27715
; LENGTH: 151
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004468.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.82
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.89
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.84
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.88
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.81
; OTHER INFORMATION: NT HIT: M18533.1, EVALUE 6.00e-80
; OTHER INFORMATION: SWISSPROT HIT: P11532, EVALUE 4.00e-08
US-09-864-761-27715

Query Match 7.4%; Score 151; DB 10; Length 151;
Best Local Similarity 100.0%; Pred. No. 7.4e-36;
Matches 151; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 382 TTTACATAGAGTTTAAATGGATCTCCAGAACTCAGAAACTGAAAGAGTTGAATGACTGGCT 441
|||||
Db 1 TTTACATAGAGTTTAAATGGATCTCCAGAACTCAGAAACTGAAAGAGTTGAATGACTGGCT 60

RESULT 6
 US-09-918-995-24084
 Sequence 24084, Application US/09918995
 Publication No. US20030073623A1
 GENERAL INFORMATION:
 APPLICANT: Hyseq, Inc.
 TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
 TITLE OF INVENTION: FROM VARIOUS CDNA LIBRARIES
 FILE REFERENCE: 20411-756
 CURRENT APPLICATION NUMBER: US/09/918,995
 CURRENT FILING DATE: 2001-07-30
 PRIOR APPLICATION NUMBER: US/09/235,076
 PRIOR FILING DATE: 1999-01-20
 NUMBER OF SEQ ID NOS: 38054
 SOFTWARE: FastSeq for Windows version 3.0
 SEQ ID NO 24084

[illegible]

RESULT 8

Db 173 ATCTAGACACAGAATCCTT 155

RESULT 15

US-09-822-846-343/c
; Sequence 343, Application US/09822846
; Publication No. US20030027139A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Agostino, Michael J.
; APPLICANT: Steininger II, Robert J.
; APPLICANT: Bowman, Michael R.
; APPLICANT: Spaulding, Vikki
; APPLICANT: Wong, Gordon G.
; APPLICANT: Clark, Hilary
; APPLICANT: Fechtel, Kim
; APPLICANT: Howes, Steven H.
; APPLICANT: Resnick, Richard J.
; APPLICANT: Gulukota, Kamalakar
; APPLICANT: Graham, James R.
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS
; FILE REFERENCE: GIN 6400
; CURRENT APPLICATION NUMBER: US/09/822,846
; CURRENT FILING DATE: 2001-03-29
; PRIOR APPLICATION NUMBER: 60/195,605
; PRIOR FILING DATE: 2000-04-06
; NUMBER OF SEQ ID NOS: 629
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 343
; LENGTH: 1381
; TYPE: DNA
; ORGANISM: Xenopus sp.
US-09-822-846-343

	Query Match	1.9%	Score 39.6;	DB 9;	Length 1381;
	Best Local Similarity	60.0%	Pred. No. 0.34;		
	Matches	66;	Conservative	0;	Mismatches 44; Indels 0; Gaps 0;
Qy	369	AAAAACAAAGCAATTTACATAGAGT	TTTAAATGGATCTCCAGATCAGAAACTGAAAAGAGT	428	
Db	1087	AAAAACAATTTAAATATTTCAAATTTTAAATTTGCTCCACCATAGATGAATAAGAGC	1028		
Qy	429	TGAATGACTGGCTAACAAAAACAGAAAGAAACAAAGGAAAAATGGAGGAA	478		
Db	1027	TTACTTAAGGAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGGAA	978		

Search completed: April 25, 2003, 23:58:31
Job time : 145.776 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 07:56:07 ; Search time 3411.5 Seconds
(without alignments)
17070.103 Million cell updates/sec

Title: US-09-845-416-14_COPY_1000_3000
Perfect score: 2001
Sequence: 1 ggcagttcattgatggagag.....cattgagccaagtgtccgga 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues
Total number of hits satisfying chosen parameters: 3813070

Minimum DB seq length: 0
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_mam:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1332	66.6	4402	6	E30220	E30220 Shortened d
2	1269.2	63.4	4402	6	E30219	E30219 Shortened d
3	776.6	38.8	3275	10	MUSDYS	M18025 Mouse dyst
4	735.4	36.8	4075	6	E30221	E30221 Shortened d
5	731.2	36.5	3747	6	E30218	E30218 Shortened d
6	677.6	33.9	3163	6	E30223	E30223 Shortened d
7	615.4	30.8	2654	5	FSCDYSTRO	M37645 Torpedo cal
8	612.2	30.6	630	9	HSDMDF1	X06179 Human fetal
9	569.	28.4	2110	9	HUMDMDX	M92650 Human Duch
10	569	28.4	2110	9	HUMDMDX	BC028720 Homo sapi
11	534	26.7	3521	5	AF339031	AF339031 Danio rer
12	528	26.4	1966	9	HSDMDA1	X06178 Human adult
13	518	25.9	1240	10	RNAPDY3	X69767 R.norvegicu
14	506.6	25.3	3161	10	MMGUTRPH	X83506 M.musculus
15	475	23.7	3073	10	AF195788	AF195788 Rattus no
16	475	23.7	3140	10	AF195787	AF195787 Rattus no
17	470.2	23.5	3499	9	HSU43519	U43519 Human dyst
18	458.8	22.9	3172	6	E30222	E30222 Shortened d
19	387.6	19.4	1961	5	XLDYSTROP	X99700 X.laavis mR
20	358.2	17.9	1195	5	AF339032	AF339032 Danio rer
21	350.6	17.5	1993	5	SCDYSTROP	X99702 S.caniculua
22	345.8	17.3	3650	3	SPAJ3356	AJ223356 Strongylo
23	334.2	16.7	1898	10	BC024140	BC024140 Mus muscu
24	265	13.2	506	10	AB011666	AB011666 Rattus no
25	264.2	13.2	1428	5	SCU43517	U43517 Scyliorhinu
26	263.6	13.2	1482	5	XLUTROPHI	X99701 X.laavis mR
27	255.4	12.8	1737	5	SCUTROPHI	X99703 S.caniculua
28	241.6	12.1	384	10	RNADYI	X65468 R.norvegicu
29	232.2	11.6	1440	5	GSU43518	U43518 Gobius sp.
30	225.2	11.3	1428	10	MMU43520	U43520 Mus musculu
31	213.2	10.7	1413	3	ASDYSTROP	X99737 Asteroidea
32	210.8	10.5	1431	3	PSDYSTROP	X99738 Pectinidae
33	203	10.1	351	9	HUMDYSTO6	M86889 H.sapiens d
34	196.8	9.8	1338	3	BLDYSTROP	X99736 B.lanceolat
35	187	9.3	525	9	AF213411	AF213411 Homo sapi
36	186.2	9.3	1500	6	AX107969	AX107969 Sequence
37	186.2	9.3	1500	6	AX107970	AX107970 Sequence
38	180.4	9.0	454	9	AF213415	AF213415 Homo sapi
39	179.8	9.0	374	10	S62620	S62620 (mdx3Cy)=dy
40	175.4	8.8	1729	3	CIDYSTRO	X99756 C.intestina
41	160.2	8.0	645	9	AF213410	AF213410 Homo sapi
42	157	7.8	300	9	S38776	S38776 Homo sapien
43	151	7.5	332	9	HUMDYSTO1	M86884 H.sapiens d
44	150.2	7.5	2962	3	DMDYDTRO	X99757 D.melanogas
45	150	7.5	234	9	S60971	S60971 dystrophin

ALIGNMENTS

RESULT 1
E30220
LOCUS E30220
DEFINITION Shortened dystrophin.
ACCESSION E30220
VERSION E30220.1 GI:13017027
KEYWORDS JP 1999318467-A/3.
SOURCE unidentified.
ORGANISM unidentified
REFERENCE 1 (bases 1 to 4402)
AUTHORS Sinichi,T.
TITLE Shortened dystrophin
JOURNAL Patent: JP 1999318467-A 3 24-NOV-1999;
SCIENCE & TECH AGENCY,NATIONAL CENTER OF NEUROLOGY AND PSYCHIATRY

COMMENT		OS	Unidentified	1915	TTTAGTGCATGGCTTT	CAGAAAAAGAGATGCAGTGAACAAGATT	CACAACA	ACTGGCTTT	1974
PN	JP 199318467-A/3	PD	24-NOV-1999	QY	780	AAAGATCAAAATGAAATGTTATCAAGTCTT	CAAAAACTGGCCGTTT	TAAAAAGCGATCTA	839
PF	08-MAY-1998	PR	JP 1998142134	Db	1975	AAAGATCAAAATGAAATGTTATCAAGT	-----	-----	2001
PI	SINICHI TAKEDA	PC	C12N15/09,A61K48/00,C12N15/00	QY	840	GAAGAAGAAAGCAATCCATGGGCAAACTGTATT	CACTCAAAACAAGATCTTCTTT	CAACA	899
CC	Strandedness: Both;	CC	Topology: Linear;	Db	2002	-----	-----	-----	2001
FH	Key	FT	Location/Qualifiers	QY	900	CTGAAGAATAAGTCAGTGACCCAGAAAGACGGAAGCATGGCTGGATAA	ACTTTTGCCCGGTGT		959
FT	source	FT	1. .4402	Db	2002	-----	-----	-----	2001
FEATURES	Location/Qualifiers	FT	/organism='Unidentified'.	QY	1020	GAACCTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGGCCAAAGCTGAGGTGATCAAG			1079
source	1. .4402	Db	/organism="unidentified"	Db	2002	-----	-----	-----	2001
BASE COUNT	1339 a 984 c 1010 g 1069 t	QY	/db_xref="taxon:32644"	QY	960	TGGGATAATTTAGTCCAAAAAACTTGAAAAAGAGTACAGCACAGACCCCTTGAAAGACTCCAG			1019
ORIGIN		Db		Db	2002	-----	-----	-----	2001
	Query Match 66.6%; Score 1332; DB 6; Length 4402;	QY		QY	1140	GTCAAGGCACCTTCGAGGAGAAATTGCGCCTCTGAAAGAGAAACGTGAGCCACGTC	AAATGAC		1199
	Best Local Similarity 83.8%; Pred. No. 0;	Db		Db	2011	GTCAAGGCACCTTCGAGGAGAAATTGCGCCTCTGAAAGAGAAACGTGAGCCACGTC	AAATGAC		2070
	Matches 1677; Conservative 0; Mismatches 0; Indels 325; Gaps 2;	QY		QY	1200	CTTGCTGCCAGCTTACCACCTTTGGGCATTCAGCTCTCACCGTATAAACCTCAGCACTCTG			1259
QY	1 GGCAGTTTCATTGATGGAGAGTGAAGTAAACCTGGACCGTTATCAAAACAGCTTTTAGAAGAA	Db		Db	2002	-----	-----	-----	2010
Db	1195 GGCAGTTTCATTGATGGAGAGTGAAGTAAACCTGGACCGTTATCAAAACAGCTTTTAGAAGAA	QY		QY	1260	GAAGACCTGAACACCCAGATGGAAGCTTTCGAGGTGGCCGTCGAGGACCGAGTCAAGCAG			1319
QY	61 GTATTATCGTGGCTTCTTCTGCTGAGGACACATTGCAAGCACAGATTGCAAGCACAGGAGATTCTTAAT	Db		Db	2131	GAAGACCTGAACACCCAGATGGAAGCTTTCGAGGTGGCCGTCGAGGACCGAGTCAAGCAG			2190
Db	1255 GTATTATCGTGGCTTCTTCTGCTGAGGACACATTGCAAGCACAGGAGAGATTCTTAAT	QY		QY	1320	CTGCATGAAGCCCCACAGGGACTTTGGTCCAGCATCTCAGCACCTTTCTTTCCACGCTCTGTC			1379
QY	121 GATGTGGAAGTGGTGAAGACACAGTTTCATACTCATGAGGGGTACATGATGGATTGACA	Db		Db	2191	CTGCATGAAGCCCCACAGGGACTTTGGTCCAGCATCTCAGCACCTTTCTTTCCACGCTCTGTC			2250
Db	1315 GATGTGGAAGTGGTGAAGACACAGTTTCATACTCATGAGGGGTACATGATGGATTGACA	QY		QY	1380	CAGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCTACTATATCAACCACCGAG			1439
QY	181 GCCCATCAGGGCCGGTTGGTAATATTCTACAATTGGGAAGTAAAGCTGATTTGGAACAGGA	Db		Db	2251	CAGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCTACTATATCAACCACCGAG			2310
Db	1375 GCCCATCAGGGCCGGTTGGTAATATTCTACAATTGGGAAGTAAAGCTGATTTGGAACAGGA	QY		QY	1440	ACTCAAAACAACCTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGAC			1499
QY	241 AAATTATCAGAAGATGAAGAAACTGAAGTACAAGAGCAGATGAATCTCCTAAATTC	Db		Db	2311	ACTCAAAACAACCTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGAC			2370
Db	1435 AAATTATCAGAAGATGAAGAAACTGAAGTACAAGAGCAGATGAATCTCCTAAATTC	QY		QY	1500	CTGAATAATGTCAGATTTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAG			1559
QY	301 TGGGAATGCCTCAGGGTAGCTAGCATGGAAGAGTTGAATGACTGGCTAACCAAAACAGAAAGAA	Db		Db	2371	CTGAATAATGTCAGATTTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAG			2430
Db	1495 TGGGAATGCCTCAGGGTAGCTAGCATGGAAGAGTTGAATGACTGGCTAACCAAAACAGAAAGAA	QY		QY	1560	GCCTTTTGCTTGGATCTCTTGAGCCCTGTGAGCTGCATGTGATGCCTTGACCTGACATAAC			1619
QY	361 GATCTCCAGAAATC-GAAACTGAAAGAGTTGAATGACTGGCTAACCAAAACAGAAAGAA	Db		Db	2431	GCCTTTTGCTTGGATCTCTTGAGCCCTGTGAGCTGCATGTGATGCCTTGACCTGACATAAC			2490
Db	1555 GATCTCCAGAAATCAGAAACTGAAAGAGTTGAATGACTGGCTAACCAAAACAGAAAGAA	QY		QY	1620	CTCAAGCAAAATGACCCAGCCCATGGATATCCTGCAGATTATTAATTTGTTGACCACTATT			1679
QY	420 ACAAGGAAATGGAGGAAGAGCCCTCTTGACCTGATCTTGAAGACCTTAAACGCCAAGTA	Db		Db	2491	CTCAAGCAAAATGACCCAGCCCATGGATATCCTGCAGATTATTAATTTGTTGACCACTATT			2550
Db	1615 ACAAGGAAATGGAGGAAGAGCCCTCTTGACCTGATCTTGAAGACCTTAAACGCCAAGTA	QY		QY	1680	TATGACCCGCTGGAGCAAGAGACACAACAATTGGTCAACGTCCTCTCTCGTGGATATG			1739
QY	480 CAACAACATAAGGTGCTTCAAGAAGATCTAGAACAAAGACAAAGTCAAGGTCAATTTCTC	Db		Db	2551	TATGACCCGCTGGAGCAAGAGACACAACAATTGGTCAACGTCCTCTCTCGTGGATATG			2610
Db	1675 CAACAACATAAGGTGCTTCAAGAAGATCTAGAACAAAGACAAAGTCAAGGTCAATTTCTC	QY		QY	1740	TGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCGTTGCTCTG			1799
QY	540 ACTCACATGGTGGTGGTAGTTGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAA	Db		Db	2611	TGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCGTTGCTCTG			2670
Db	1735 ACTCACATGGTGGTGGTAGTTGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAA	QY		QY	1800	TCTTTAAAACTGGCATCATTTCCCTGTGTAAAGCACATTTTGGAAAGACAAAGTACAGATAC			1859
QY	600 GAACAACCTTAAGGTATTGGGAGATCGATGGGCAACATCTGTAGATGGACAGAACCCGC	Db		Db	2671	TCTTTAAAACTGGCATCATTTCCCTGTGTAAAGCACATTTTGGAAAGACAAAGTACAGATAC			


```
QY 1860 CTTTTCAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAGCGCAGGCTGGGCCTCCTT 1919
|||||
Db 2731 CTTTTCAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAGCGCAGGCTGGGCCTCCTT 2790
|||||
QY 1920 CTGCATGATTCATCCAAATTCCAAGACAGTTGGGTGAAGTTGCATCCTTTGGGGGCAGT 1979
|||||
Db 2791 CTGCATGATTCATCCAAATTCCAAGACAGTTGGGTGAAGTTGCATCCTTTGGGGGCAGT 2850
|||||
QY 1980 AACATTGAGCCAAGTGTCGGGA 2001
|||||
Db 2851 AACATTGAGCCAAGTGTCGGGA 2872
|||||

RESULT 2
E30219
LOCUS E30219 4402 bp DNA linear PAT 18-JUN-2001
DEFINITION Shortened dystrophin.
ACCESSION E30219
VERSION E30219.1 GI:13017026
KEYWORDS JP 1999318467-A/2.
SOURCE unidentified.
ORGANISM unidentified
unclassified.
REFERENCE 1 (bases 1 to 4402)
AUTHORS Sinichi,T.
TITLE Shortened dystrophin
JOURNAL Patient: JP 1999318467-A 2 24-NOV-1999;
SCIENCE & TECH AGENCY,NATIONAL CENTER OF NEUROLOGY AND PSYCHIATRY
COMMENT OS Unidentified
PN JP 1999318467-A/2
PD 24-NOV-1999
PF 08-MAY-1998 JP 1998142134
PR
PI SINICHI TAKEDA
PC C12N15/09,A61K48/00,C12N15/00
CC Strandedness: Both;
CC Topology: Linear;
FH key Location/Qualifiers
FT source 1..4402
FT /organism='Unidentified'.
FEATURES
source Location/Qualifiers
1..4402
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 1329 a 1000 c 1019 g 1054 t
ORIGIN
Query Match 63.4%; Score 1269.2; DB 6; Length 4402;
Best Local Similarity 82.1%; Pred. No. 4.1e-309;
Matches 1644; Conservative 0; Mismatches 33; Indels 325; Gaps 3;

QY 1 GGCAGTTTCATTGATGGAGAGTGAAGTAAACCTGGACCGTTATCAACACAGCTTTAGAAGAA 60
|||||
Db 1195 GGCAGTTTCATTGATGGAGAGTGAAGTAAACCTGGACCGTTATCAACACAGCTTTAGAAGAA 1254
|||||
QY 61 GTATTATCGTGGCTTCTTCTGCTGAGGACACATTGCAAGCACACAGGAGAGATTTCTAAT 120
|||||
Db 1255 GTATTATCGTGGCTTCTTCTGCTGAGGACACATTGCAAGCACACAGGAGAGATTTCTAAT 1314
|||||
QY 121 GATGTGGAAGTGTGAAAGACCAGTTTCATACCTCATGAGGGTACATGATGATTGACA 180
|||||
Db 1315 GATGTGGAAGTGTGAAAGACCAGTTTCATACCTCATGAGGGTACATGATGATTGACA 1374
|||||
QY 181 GCCCATCAGGGCGGGTGGTAATATCTACAATTGGGAAGTAAGCTGATTGGAACAGGA 240
|||||
Db 1375 GCCCATCAGGGCGGGTGGTAATATCTACAATTGGGAAGTAAGCTGATTGGAACAGGA 1434
|||||
QY 241 AAATTATCAGAAGATGAAGAAACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAAGA 300
|||||
Db 1435 AAATTATCAGAAGATGAAGAAACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAAGA 1494
|||||
QY 301 TGGGAATGCCTCAGGGTAGCTAGCATGGAAAAACAAGCAATTTACATAGAGTTTAAATG 360
|||||
```

```
Db 1495 TGGGAATGCCTCAGGCTAGCTAGCATGGAAAAACAAGCAATTTACATAGAGTTTAATG 1554
|||||
QY 361 GATCTCCAGAAATC-GAAACTGAAAGAGTTGAATGACTGGCTGAACAAAAACAGAAAGA 419
|||||
Db 1555 GATCTCCAGAAATCAGAAACTGAAAGAGTTGAATGACTGGCTGAACAAAAACAGAAAGA 1614
|||||
QY 420 ACAAGGAAATGAGGAAGAGCCTCTTGGACCTGATCTTGAAGACCTAAACGCCAAGTA 479
|||||
Db 1615 ACAAGGAAATGAGGAAGAGCCTCTTGGACCTGATCTTGAAGACCTAAACGCCAAGTA 1674
|||||
QY 480 CAACAACATAAGGTGCTTCAAGAAGATCTAGAACAAAGACAGTCAAGGTCAATTTCTCTC 539
|||||
Db 1675 CAACAACATAAGGTGCTTCAAGAAGATCTAGAACAAAGACAGTCAAGGTCAATTTCTCTC 1734
|||||
QY 540 ACTCACATGGTGGTGGTAGTTGATGAATCTAGTGGAGATCAGCAACTGCTGCTTTGGAA 599
|||||
Db 1735 ACTCACATGGTGGTGGTAGTTGATGAATCTAGTGGAGATCAGCAACTGCTGCTTTGGAA 1794
|||||
QY 600 GAACAACCTTAAGGTATTGGGAGATCGATGGGCAAAACATCTGTAGATGGACAGAAGACCGC 659
|||||
Db 1795 GAACAACCTTAAGGAGGTCAATACTAGTGGGAAAAA----- 1829
|||||
QY 660 TGGGTTCTTTTACAAGACATCCTTCTCAAAATGGCAACGCTCTTACTGAAGAACAGTGCCTT 719
|||||
Db 1830 ----- 1829
|||||
QY 720 TTTAGTGCATGGCTTTCAGAAAAAAGAGATGCAGTGAACAAGATTTCACACAACCTGGCTTT 779
|||||
Db 1830 ----- 1829
|||||
QY 780 AAAGATCAAAATGAAATGTTATCAAGTCTTCAAAAACCTGGCCGTTTAAAGCGGATCTA 839
|||||
Db 1830 ----- 1829
|||||
QY 840 GAAAAGAAAAAGCAATCCATGGGCAAACTGTATTACTCAACAAGATCTTCTTTCAACA 899
|||||
Db 1830 ----- 1829
|||||
QY 900 CTGAAGAATAAGTCAGTGACCCAGAAAGACGGAAAGCATGGCTGGATAAATTTGCCCGGTGT 959
|||||
Db 1830 -----ATTGAACCTGCACCTCCGCTGACTGGCAGAGAAA----- 1862
|||||
QY 960 TGGGATAATTAGTCCAAAAACTTGAAAAGAGTACAGCACAGACCCCTTGAAAAGACTCCAG 1019
|||||
Db 1863 -----AATAGATGAGACCCCTTGAAAAGACTCCAG 1890
|||||
QY 1020 GAACCTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGGCCCAAGCTGAGGTGATCAAG 1079
|||||
Db 1891 GAACCTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGGCCCAAGCTGAGGTGATCAAG 1950
|||||
QY 1080 GGATCCTGGCAGCCCGTGGCGGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGAGAAA 1139
|||||
Db 1951 GGATCCTGGCAGCCCGTGGCGGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGAGAAA 2010
|||||
QY 1140 GTCAAGGCACCTTCGAGGAGAAAAATTGCGCCTCTGAAAAGAGAACGTTGAGCCACGTCAATGAC 1199
|||||
Db 2011 GTCAAGGCACCTTCGAGGAGAAAAATTGCGCCTCTGAAAAGAGAACGTTGAGCCACGTCAATGAC 2070
|||||
QY 1200 CTTGCTCGCCAGCTTACCACCTTTGGGCATTTGGGCATTTAGCTCTCACCGTATAACCTCAGCACTGTG 1259
|||||
Db 2071 CTTGCTCGCCAGCTTACCACCTTTGGGCATTTAGCTCTCACCGTATAACCTCAGCACTGTG 2130
|||||
QY 1260 GAAGACCTGAACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGGAGTCAGGCAG 1319
|||||
Db 2131 GAAGACCTGAACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGGAGTCAGGCAG 2190
|||||
QY 1320 CTGCATGAAGCCCAACAGGGACTTTGGTCCAGCATCTCAGCACTTCTTTCCACGTCTGTC 1379
|||||
Db 2191 CTGCATGAAGCCCAACAGGGACTTTGGTCCAGCATCTCAGCACTTCTTTCCACGTCTGTC 2250
|||||
QY 1380 CAGGTCCCTGGGAGAGAGCCCATCTCGCCCAACAAAGCAATTTACATAGAGTTTAAATG 1439
|||||
```


||||| TTTAGTACATGGCTTTCAGAAAAAGAGATGCAATGAGAACATTCAGACAAAGTGGCTTT 881
Db 822
QY 780 AAAGATCAAAATGAAATGTTATCAAGTCTTCAAAAGTGGCGGTTTAAAGCGGATCTA 839
Db 882 AAAGATCAAAATGAAATGATGTCAAGTCTTCAAAATATCTACTTTAAATAGATCTA 941
QY 840 GAAAGAAAAAGCAATCCATGGGCAAACTGTATTCACCTCAAAAGATCTTCTTTCAACA 899
Db 942 GAAAGAAAAAGCAACCATGGAAAGTCAAGTCAATCAAGATCTACTTTTCGGCA 1001
QY 900 CTGAGAATAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 959
Db 1002 CTGAAAAATAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGT 1061
QY 960 TGGGATAATTTAGTCCAAAAAACTTGAAGAGAGTACAGCACAGA 1002
Db 1062 TGGGACAATTTAACCCAAAAAACTTGAAGAGAGTTCAGCACAAA 1104
RESULT 4
E30221
LOCUS E30221 Shortened dystrophin. 4075 bp DNA linear PAT 18-JUN-2001
DEFINITION E30221
ACCESSION E30221
VERSION E30221.1 GI:13017028
KEYWORDS JP 199318467-A/4.
SOURCE unidentifed.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 4075)
AUTHORS Sinichi,T.
TITLE Shortened dystrophin
JOURNAL SCIENCE & TECH AGENCY,NATIONAL CENTER OF NEUROLOGY AND PSYCHIATRY
COMMENT OS Unidentifed
PN JP 199318467-A/4
PD 24-NOV-1999
PF 08-MAY-1998 JP 1998142134
PR
PI SINICHI TAKEDA
PC C12N15/09,A61K48/00,C12N15/00
CC Strandedness: Both;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..4075
/organism='Unidentified'.
FEATURES
source
1..4075
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 1237 a 910 c 936 g 992 t
ORIGIN
Query Match 36.8%; Score 735.4; DB 6; Length 4075;
Best Local Similarity 99.2%; Pred. No. 1.9e-174;
Matches 739; Conservative 0; Mismatches 6; Indels 0; Gaps 0;
QY 1257 CTGGAAGACCTGAACACACAGATGGAAGCTTCTGCAGGTGGCGCTCGAGGACCGAGTCAGG 1316
Db 1801 CTTAAGGTATTGAACACACAGATGGAAGCTTCTGCAGGTGGCGCTCGAGGACCGAGTCAGG 1860
QY 1317 CAGCTGCATGAAGCCACAGGGACTTTGGTCCAGCATCTCAGCACTTCTTCTCCACGTCT 1376
Db 1861 CAGCTGCATGAAGCCACAGGGACTTTGGTCCAGCATCTCAGCACTTCTTCTCCACGTCT 1920
QY 1377 GTCCAGGTCCTGGGAGAGAGCCATCTCGCCAAACAAAGTCCCTACTATATCAACCAC 1436
Db 1921 GTCCAGGTCCTGGGAGAGAGCCATCTCGCCAAACAAAGTCCCTACTATATCAACCAC 1980
QY 1437 GAGACTCAAAACAACCTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCT 1496
Db 1981 GAGACTCAAAACAACCTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCT 2040

QY 1497 GACCTGAATAATGTGAGATTTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAG 1556
Db 2041 GACCTGAATAATGTGAGATTTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAG 2100
QY 1557 AAGSCCCTTTGCTTGGATCTCTTGGCCCTGTGAGCCTGTGAGTGCATGTGATGCCTTGGACCAAC 1616
Db 2101 AAGSCCCTTTGCTTGGATCTCTTGGCCCTGTGAGCCTGTGAGTGCATGTGATGCCTTGGACCAAC 2160
QY 1617 AACCTCAAGCAAAATGACCAAGCCCATGGATATCTCTGACAGATTATTAATTGTTTGACCACT 1676
Db 2161 AACCTCAAGCAAAATGACCAAGCCCATGGATATCTCTGACAGATTATTAATTGTTTGACCACT 2220
QY 1677 ATTTATGACCCCTGGAGCAAGAGCACAAATTTGGTCAACGTCCCTCTCTGCGTGGAT 1736
Db 2221 ATTTATGACCCCTGGAGCAAGAGCACAAATTTGGTCAACGTCCCTCTCTGCGTGGAT 2280
QY 1737 ATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGAGGATCCGTGTC 1796
Db 2281 ATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGAGGATCCGTGTC 2340
QY 1797 CTGTCTTTTAAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAAGTACAGA 1856
Db 2341 CTGTCTTTTAAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAAGTACAGA 2400
QY 1857 TACCTTTTCAAGCAAGTGGCAAGTTCACAGGATTTTGTGACCAAGCGGCTGGGCGCTC 1916
Db 2401 TACCTTTTCAAGCAAGTGGCAAGTTCACAGGATTTTGTGACCAAGCGGCTGGGCGCTC 2460
QY 1917 CTTCTGCATGATTCTATCCAAATTCACAGACAGTTGGGTGAAGTTGCATCCTTTGGGGGC 1976
Db 2461 CTTCTGCATGATTCTATCCAAATTCACAGACAGTTGGGTGAAGTTGCATCCTTTGGGGGC 2520
QY 1977 AGTAACATTGAGCCCAAGTGTCCGGA 2001
Db 2521 AGTAACATTGAGCCCAAGTGTCCGGA 2545

RESULT 5
E30218
LOCUS E30218 Shortened dystrophin. 3747 bp DNA linear PAT 18-JUN-2001
DEFINITION E30218
ACCESSION E30218
VERSION E30218.1 GI:13017025
KEYWORDS JP 199318467-A/1.
SOURCE unidentifed.
ORGANISM unclassified.
REFERENCE 1 (bases 1 to 3747)
AUTHORS Sinichi,T.
TITLE Shortened dystrophin
JOURNAL SCIENCE & TECH AGENCY,NATIONAL CENTER OF NEUROLOGY AND PSYCHIATRY
COMMENT OS Unidentifed
PN JP 199318467-A/1
PD 24-NOV-1999
PF 08-MAY-1998 JP 1998142134
PR
PI SINICHI TAKEDA
PC C12N15/09,A61K48/00,C12N15/00
CC Strandedness: Both;
CC Topology: Linear;
FH Key Location/Qualifiers
FT source 1..3747
/organism='Unidentified'.
FEATURES
source
1..3747
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 1112 a 853 c 859 g 923 t
ORIGIN
Query Match 36.5%; Score 731.2; DB 6; Length 3747;

	Best Local Similarity Matches 736;	Conservative	98.9%;	Pred. No. 2.le-173;	Mismatches 0;	Indels 8;	Gaps 0;
QY	1258	TGGAAGACCTGAACACACCAGATGGAAGCTTCTGCAGGTGGCCGTGCAGGACCGAGTCAGGC	1317				
Dbb	1475	TGAATCTCCTAAATTCAAGATGGAAGCTTCTGCAGGTGGCCGTGCAGGACCGAGTCAGGC	1534				
QY	1318	AGCTGCATGAAGCCACACAGGGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGCTG	1377				
Dbb	1535	AGCTGCATGAAGCCACACAGGGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGCTG	1594				
QY	1378	TCCAGGGTCCCTGGGAGAGAGCCATCTGCCAAACAAAGTGCCCTACTATATCAACCCACG	1437				
Dbb	1595	TCCAGGGTCCCTGGGAGAGAGCCATCTGCCAAACAAAGTGCCCTACTATATCAACCCACG	1654				
QY	1438	AGACTCAAAACAACCTTGTCTGGGACCATCCCAAATGACAGAGCTCTACCACTCTTTAGCTG	1497				
Dbb	1655	AGACTCAAAACAACCTTGTCTGGGACCATCCCAAATGACAGAGCTCTACCACTCTTTAGCTG	1714				
QY	1498	ACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGA	1557				
Dbb	1715	ACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGA	1774				
QY	1558	AGGCCCTTTGCTTGGATCTCTTGAGCCTGTTCAGCTGCATGTGATGCCCTTGACCAGCACA	1617				
Dbb	1775	AGGCCCTTTGCTTGGATCTCTTGAGCCTGTTCAGCTGCATGTGATGCCCTTGACCAGCACA	1834				
QY	1618	ACCTCAAGCAAAATGACCAGGCCCATGGATATCCTGCAGATTATTAATGTTTGACCACTA	1677				
Dbb	1835	ACCTCAAGCAAAATGACCAGGCCCATGGATATCCTGCAGATTATTAATGTTTGACCACTA	1894				
QY	1678	TTTATGACCGCCTGGAGCAAGAGCACAACAATTTGGTCAACGTCCCTCTCTGCGTGGATA	1737				
Dbb	1895	TTTATGACCGCCTGGAGCAAGAGCACAACAATTTGGTCAACGTCCCTCTCTGCGTGGATA	1954				
QY	1738	TGTGCTGAACCTGSGCTGCTGAATGTTTATGATACGGGACGAAACAGGGAGGATCCGTGTCC	1797				
Dbb	1955	TGTGCTGAACCTGSGCTGCTGAATGTTTATGATACGGGACGAAACAGGGAGGATCCGTGTCC	2014				
QY	1798	TGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAACACAAAGTACAGAT	1857				
Dbb	2015	TGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAACACAAAGTACAGAT	2074				
QY	1858	ACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAGCGAGGCTGGGCCCTCC	1917				
Dbb	2075	ACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAGCGAGGCTGGGCCCTCC	2134				
QY	1918	TTCTGCATGATTCTATCCAAATTCGAAGAGAGATTTGGTGAAGTTGCATCCTTTGGGGGCA	1977				
Dbb	2135	TTCTGCATGATTCTATCCAAATTCGAAGAGAGATTTGGTGAAGTTGCATCCTTTGGGGGCA	2194				
QY	1978	GTAACATTGAGCCCAAGTGTCCGGA	2001				
Dbb	2195	GTAACATTGAGCCCAAGTGTCCGGA	2218				

RESULT	6
E30223	
LOCUS	E30223 Shortened dystrophin.
DEFINITION	E30223
ACCESSION	E30223.1 GI:13017030
VERSION	JP 1999318467-A/6.
KEYWORDS	unidentified.
SOURCE	unidentified
ORGANISM	unclassified.
REFERENCE	1 (bases 1 to 3163)
AUTHORS	Sinichi,T.
TITLE	Shortened dystrophin
JOURNAL	Patent: JP 1999318467-A 6 24-NOV-1999;
	SCIENCE & TECH AGENCY,NATIONAL CENTER OF NEUROLOGY AND PSYCHIATRY
COMMENT	OS Unidentified
	PN JP 1999318467-A/6

PD	24-NOV-1999
PF	08-MAY-1998 JP 1998142134
PR	
PI	SINICHI TAKEDA
PC	C12N15/09,A61K48/00,C12N15/00
CC	Strandedness: Both;
CC	Topology: Linear;
FH	Key
FT	source
FT	1..3163
FEATURES	Location/Qualifiers
source	/organism='Unidentified'.
BASE COUNT	930 a 731 c 720 g 782 t
ORIGIN	
	Query Match 33.9%; Score 677.6; DB 6; Length 3163;
	Best Local Similarity 98.7%; Pred. No. 7.1e-160;
	Matches 683; Conservative 0; Mismatches 9; Indels 0; Gaps 0;
QY	1310 AGTCAGGCAGCTGCAATGAAGCCCACAGGGACTTTGGTCCAGCATCTCAGCACATTTCTTTC 1369
Dd	942 AGCCATCCAGGAAGTGGAAGCCCCACAGGGACTTTGGTCCAGCATCTCAGCACATTTCTTTC 1001
QY	1370 CACGTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCAAACAAGTGCCCTACTATAT 1429
Dd	1002 CACGTCTGTCCAGGGTCCCTGGGAGAGAGGCCATCTCGCCAACAAGTGCCCTACTATAT 1061
QY	1430 CAACCACGAGACTCAAAACAACCTTGCTGGGACCATCCCAAATGACAGAGCTCTACCAGTC 1489
Dd	1062 CAACCACGAGACTCAAAACAACCTTGCTGGGACCATCCCAAATGACAGAGCTCTACCAGTC 1121
QY	1490 TTTAGCTGACCTGAATAATGTGAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAG 1549
Dd	1122 TTTAGCTGACCTGAATAATGTGAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAG 1181
QY	1550 ACTGCAGAAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTGAGCTGCATGTGATGCCCTTGA 1609
Dd	1182 ACTGCAGAAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTGAGCTGCATGTGATGCCCTTGA 1241
QY	1610 CCAGCACAAACCTCAAGCAAATGACAGACCCCATGGATATCCTGCAGATTATTAATTGTTT 1669
Dd	1242 CCAGCACAAACCTCAAGCAAATGACAGACCCCATGGATATCCTGCAGATTATTAATTGTTT 1301
QY	1670 GACCACTATTTATGACCGCCTGGAGCAAGAGACACAACAATTTGGTCAACGTCCCTCTCTG 1729
Dd	1302 GACCACTATTTATGACCGCCTGGAGCAAGAGACACAACAATTTGGTCAACGTCCCTCTCTG 1361
QY	1730 CGTGGATATGTGCTGAACCTGGCTGGCTGAATGTTTATGATACGGGACGACAGGGAGGAT 1789
Dd	1362 CGTGGATATGTGCTGAACCTGGCTGGCTGAATGTTTATGATACGGGACGACAGGGAGGAT 1421
QY	1790 CCGTGTCTGTCTTTTAAAACCTGGCATCATTTCCCTGTGTAAAGCACATTTTGGAAAGACAA 1849
Dd	1422 CCGTGTCTGTCTTTTAAAACCTGGCATCATTTCCCTGTGTAAAGCACATTTTGGAAAGACAA 1481
QY	1850 GTACAGATACCTTTTCAAGCAAGTGGCAAGTTCACAGGATTTTGTGACCCAGCGCAGGCT 1909
Dd	1482 GTACAGATACCTTTTCAAGCAAGTGGCAAGTTCACAGGATTTTGTGACCCAGCGCAGGCT 1541
QY	1910 GGGCCTCCTTCTGTCATGATTCTATCCAAATTCGAAGACAGTTGGGTGAAGTTGCATCCTT 1969
Dd	1542 GGGCCTCCTTCTGTCATGATTCTATCCAAATTCGAAGACAGTTGGGTGAAGTTGCATCCTT 1601
QY	1970 TGGGGGCAGTAACATTGAGCCAAAGTGTCGGGA 2001
Dd	1602 TGGGGGCAGTAACATTGAGCCAAAGTGTCGGGA 1633

RESULT 7
FSCDYSTRO
LOCUS

VRT 08-SEP-1993

DEFINITION Torpedo californica dystrophin mRNA, 3' end.
ACCESSION M37645
VERSION M37645.1 GI:397970
KEYWORDS dystrophin.
SOURCE Torpedo californica, cDNA to mRNA.
ORGANISM Torpedo californica
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
Elasmobranchii; Squalea; Hypnosqualea; Pristiogaster; Batoidae;
Torpediniformes; Torpedinoidei; Torpedinidae; Torpedo.
REFERENCE 1 (bases 1 to 2654)
AUTHORS Yeadon,J.E., Lin,H., Dyer,S.M. and Burden,S.J.
TITLE Dystrophin is a component of the subsynaptic membrane
JOURNAL J. Cell Biol. 115 (4), 1069-1076 (1991)
MEDLINE 92064638
PUBMED 1720119
REFERENCE 2 (bases 1 to 2654)
AUTHORS Ravin,A.J., Dyer,S.M., Yeadon,J.E. and Burden,S.J.
TITLE Multiple dystrophin isoforms are associated with the postsynaptic
membrane of Torpedo electric organ
JOURNAL J. Physiol. (Paris) 85 (3), 131-133 (1991)
MEDLINE 92291902
PUBMED 1818109
COMMENT On Sep 8, 1993 this sequence version replaced gi:213232.
Draft entry and computer-readable sequence for [Unpublished (1990)]
kindly submitted
by S.J.Burden, 06-AUG-1990.
Biology Dept, 16-820
MIT
Cambridge, MA 02139.

FEATURES
source Location/Qualifiers
1..2654
/organism="torpedo californica"
/db_xref="taxon:7787"
1..2654
/gene="dystrophin"
<1..2654
/gene="dystrophin"
/codon_start=3
/product="dystrophin"
/protein_id="AAC38002.1"
/db_xref="GI:397971"
/translation="SGEQWKRLOISLQDFTLWMLNKLDELRRQPIGGDAPTVCQND
VHRIKRELKAKEPVVMSALDVLHFLADPAIRAPESLTGPREKIPENIQNVAKRI
RKYAEVKEVMDKLSNRSDWQKRIIDEALKRLLEQDSDELNLKLRQAEAIKDTWQ
VGDLIDSQDHIKVKVFAEIAIPMKENVTHMNDLASQFTPPDIQLSPYNLNQLEDL
NTRKLLQVSIIDELLKQLHEHRDFGPTSQHFLSTVQGPWERAISPKNVPYINHQ
QTTCWDHPKMTLEYQLADLNNVRFSAIRYAMKRLRLQKALCLDLNLSACEAFDQ
NLKQDLQDLILEIINCLTSIYDRLEQHSNLVNVPLCVDMCLNWLNNVDTGRTGKI
RVLSFKSGIMSMCKAHLEDKRYLFKQVAPSGFCQRRIGLILHEAIIQIROLGEVA
SFGSGNIEPSPVRSQCFQANNKPEIEAALFDMRLPEQSLVWMPVLRVAAAEAKHQ
AKCNICKEPIIGFRYSLKHFNDVQCFFSGRTAKGKMHYPMVEYCTPTTSGED
VRDFAKVLKNKFTKRYFAKHPRMGYLPVQTVLEGNLETPTVLINFPVDYEPASSP
QLSHDDTHSRIEHLRLAEMENRNGSYLNDSPNESIDDEHLILQHYCQSLNQESP
LSQPRSPAQILISLESEERGELEIRILADLEENRLQSEYEKLKQHDHKGSLPLSP
PEMPFISQSPRDAELIAEAKLLRQKGRLEARMQILEDHKNQLESOLHRLRLLEQ
QAEVRNGTSVSSPSTSSQSDSSQPVLLHGVGSQTSGLGEDDLLSAPQSTSELED
VMEQLSSFPSPSQRRTDTPGKKMKEVPM"
BASE COUNT 773 a 618 c 646 g 617 t
ORIGIN

Query Match 30.8%; Score 615.4; DB 5; Length 2654;
Best Local Similarity 75.3%; Pred. No. 3.5e-144;
Matches 766; Conservative 0; Mismatches 251; Indels 0; Gaps 0;
QY 985 AAAAGAGTACAGCAGACACCCCTTGAAAGACTCCAGGAACTCAAGAGGCCACGGATGAGC 1044
Db 373 AGAAGCGTATAGATGAGGCCTTGAAGAGACTACTGGAAATGCAAGATTCAATGGATGAAT 432
QY 1045 TGGACCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGATTCCTGGCAGCCCGTGGCGGATC 1104
Db 433 TGAACCTCAATGAGACAGGCTGAAGCTATCAAAGATACATGGCAACCTGTCGGGGATC 492
QY 1105 TCCTCATTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAATTG 1164

Db 493 TACTAATAGACTCATTGCAGGATCACATTGAAAAAGTCRAGGTTTTTCGAGCAGAAATTG 552
QY 1165 CGCCTCTGAAAGAGAACGTGAGCCACGTCAATGACCTTCTGTCGCCAGCTTACCACCTTGG 1224
Db 553 CTCCCATGAAGGAAATGTGACTCACATGAATGATCTCGCTTCCAGTTCACACACCCTG 612
QY 1225 GCATTCACTCTCACCGTATACCTCAGCAGCTCTGGAAGACCTGAACACACAGATGGAAGC 1284
Db 613 ATATCCAATTATCCCCGTACAAATCTAAACCAGTTGGAGGATCTGAACACACACGGTGGAAC 672
QY 1285 TTCTGCAGGTGGCCGTCGAGGACCGAGTCAAGGAGTGCATGAAGCCACACAGGACTTTG 1344
Db 673 TCCTGCAGGTGTCTATAGATGAGCTTCTGAAGCAGCTGATGAGGCTCATAGAGATTTTG 732
QY 1345 GTCCAGCATCTCAGCAGCTTTCTTTCCACGTCTGCCAGGGTCCCTGGGAGAGGACCATCT 1404
Db 733 GACCAACATCCAGCAGCTTCTCTATCAACTTCTGCCAAGGTCTCTGGAGCGGACATAT 792
QY 1405 CGCCAAACAAAGTGCCCTACTATATCAACACAGAGACTCAAAACAACTTGTGGGACCATC 1464
Db 793 CACCAACAAAGTTCCCTACTATCACTAATTAACCAACCAACTCAGACCACCTGCTGGGATCAC 852
QY 1465 CCAAAATGACAGAGCTCTACCCAGCTCTTTAGCTGACCTGAATAATGTGAGATTCTCAGCTT 1524
Db 853 CCAAGATGACGGAGCTCTATCAGTCACTAGCCGATCTAAATAATGTGAGGTTTCTGCTGCT 912
QY 1525 ATAGGACTGCCATGAAACTCCGAGAGACTGCGAAGAGGCGCTTTGCTTGGATCTCTTGAGCC 1584
Db 913 ATAGAACTGCCATGAAACTCGGAAGACTGCAGAAAGCTCTTTGCTTGGATCTCTCTCC 972
QY 1585 TGTCACTGCTGATGATGCTTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1644
Db 973 TGCCAAAGTCTGTGAGGCGCTTGTATCAGCAGCAACCTGAAACAGAGCACTCAACTGTTGG 1032
QY 1645 ATATCCTGCAGATTATTAATTGTTGACCACTATTTATGACCGCTGGAGCAAGAGCACA 1704
Db 1033 ATATTTAGAAATAATTAACCTGCTGACTTCAATCTATGATCGGCTGGAGCAGGAAACACA 1092
QY 1705 ACAATTTGGTCAACGTCCTCTCTCGTGGGATATGTCTGAACTGGCTGCTGAATGTTT 1764
Db 1093 GCAACCTTGTCAACGTCCTCTTGTGTGGACATGTGTTTAAACTGGCTGCTCAACGCTCT 1152
QY 1765 ATGATACGGGACGAAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1824
Db 1153 ATGACACTGGTCCGAAACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1212
QY 1825 TGTGTAAGCAGCATTGGAACAGCAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAA 1884
Db 1213 TGTGCAAAAGCACACCTGGAGGACAAATACCGTTATCTGTTCAAGCAAGTGGCGAGTCCCA 1272
QY 1885 CAGGATTTTGTGACCAAGCGCAGGCTGGGCGCTCTCTGTCATGATTTCTATCCAAATCCAA 1944
Db 1273 CTGGATTCTGTGATCAGCGCAGGCTTGGGCTTCTACTGTCATGAGGCGATTCAGATTCCAC 1332
QY 1945 GACAGTTGGGTGAAGTTGCATCTTTGGGGGCGAGTAAACATTGAGCCAAAGTGTCCGGA 2001
Db 1333 GCCAGCTGGGAGAGGTTGCATCTCTTTGGAGGCGAGCAACATAGAACCCAGTGTTCGCA 1389

RESULT 8
HSDMDFI
LOCUS Human fetal mRNA fragment of DMD gene (DMD= Duchenne muscular dystrophy), 630 bp mRNA linear PRI 28-MAR-1991
DEFINITION
ACCESSION X06179 M22819
VERSION X06179.1 GI:30839
KEYWORDS Duchenne muscular dystrophy.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 630)

AUTHORS Cross,G.S., Speer,A., Rosenthal,A., Forrest,S.M., Smith,T.J.,
Edwards,Y., Flint,T., Hill,D. and Davies,K.E.
TITLE Deletions of fetal and adult muscle CDNA in Duchenne and Becker
muscular dystrophy patients
JOURNAL EMBO J. 6 (11), 3277-3283 (1987)
MEDLINE 88111512
PUBMED 3428261
COMMENT Data kindly reviewed (12.OCT.1988) by DAVIES K.E.
FEATURES Location/Qualifiers
source
1. .630
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="Cf16."
/dev_stage="fetus"
1. .630
/gene="DMD"
<1. .>630
/gene="DMD"
/codon_start=2
/protein_id="CAA29545.1"
/db_xref="GI:30840"
/translation="SKLIGTKLSEDETEVEQEMNLLNSRWELRVASMEKQSNLHR
VLMDLQNKIKELNDWLTKEERTRKMEEPGLPDLEDKRQVQKHVKVLEDLEQEQV
RVNSLTHMVVVVDESSEGDHATAALEEQKLVGDRWANICRWTEDRVLLQDTLLKWQR
LTEEQCLFSAWLSEKEDAVNKIHTTGFKDQNEMLSSLOKLAVLKADLKK"
BASE COUNT 241 a 107 c 142 g 140 t
ORIGIN
Query Match 30.6%; Score 612.2; DB 9; Length 630;
Best Local Similarity 99.4%; Pred. No. 2.2e-143;
Matches 625; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 220 AGTAAGCTGATTGGAACAGGAAATATATCAGAAGATGAAGAACTGAAGTACAAGAGCAG 279
Db 2 AGTAAGCTGATTGGAACAGGAAATATATCAGAAGATGAAGAACTGAAGTACAAGAGCAG 61
QY 280 ATGAATCTCTAAATTCAGAGATGGAATGCCTCAGGCTAGCTAGCATGGAAAAACAAAGC 339
Db 62 ATGAATCTCTAAATTCAGAGATGGAATGCCTCAGGCTAGCTAGCATGGAAAAACAAAGC 121
QY 340 AATTACATAGAGTTTAAATGGATCTCCAGAAATC-GAAACTGAAAGAGATTGAATGACTGG 398
Db 122 AATTACATAGAGTTTAAATGGATCTCCAGAAATCAGAAACTGAAAGAGATTGAATGACTGG 181
QY 399 CTAACAAAAACAGAAAGAAACAAAGAAATGGAGGAAGAGCCCTCTTGGACCTGATCTT 458
Db 182 CTAACAAAAACAGAAAGAAACAAAGAAATGGAGGAAGAGCCCTCTTGGACCTGATCTT 241
QY 459 GAAGACCTAAACGCCAAGTACAACAAATAGGTGCTTCAAGAAGATCTAGAACAAAGAA 518
Db 242 GAAGACCTAAACGCCAAGTACAACAAATAGGTGCTTCAAGAAGATCTAGAACAAAGAA 301
QY 519 CAAGTCAGGGTCAATCTCTCACTCACATGGTGGTGGTAGTTGATGAATCTAGTGGAGAT 578
Db 302 CAAGTCAGGGTCAATCTCTCACTCACATGGTGGTGGTAGTTGATGAATCTAGTGGAGAT 361
QY 579 CACGCAACTGCTGCTTTTGGAAAGAACAACTTAAAGTATTGGGAGATCGATGGCAACATC 638
Db 362 CACGCAACTGCTGCTTTTGGAAAGAACAACTTAAAGTATTGGGAGATCGATGGCAACATC 421
QY 639 TGATAGTGACAGAAAGACCGCTGGTCTTTTCAAGACATCCTTCTCAAAATGGCAACGT 698
Db 422 TGATAGTGACAGAAAGACCGCTGGTCTTTTCAAGACATCCTTCTCAAAATGGCAACGT 481
QY 699 CTTACTGAAGAACAGTGCCTTTTGTAGTGCATGGCTTTTCAAGAAAAAGAGATGCAGTGAAC 758
Db 482 CTTACTGAAGAACAGTGCCTTTTGTAGTGCATGGCTTTTCAAGAAAAAGAGATGCAGTGAAC 541
QY 759 AAGATTACACAACTGGCTTTTAAAGATCAAAATGAATGTTATCAAGTCTTCAAAAAACTG 818
Db 542 AAGATTACACAACTGGCTTTTAAAGATCAAAATGAATGTTATCAAGTCTTCAAAAAACTG 601

QY 819 GCCGTTTAAAAAGCGGATCTAGAAAAAGAA 847
Db 602 GCCGTTTAAAAAGCGGATCTAAAAAAGAA 630
RESULT 9
HUMDMXX
LOCUS Human Duchenne muscular dystrophy (DMD) mRNA linear PRI 07-NOV-1994
DEFINITION Duchenne muscular dystrophy (DMD) mRNA, complete cds.
ACCESSION M92650
VERSION M92650.1 GI:181598
KEYWORDS Duchenne muscular dystrophy protein.
SOURCE Homo sapiens brain CDNA to mRNA.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2110)
AUTHORS Lederfein,D., Levy,Z., Augier,N., Mornet,D., Morris,G., Fuchs,O.,
Yaffe,D. and Nudel,U.
TITLE A 71-kilodalton protein is a major product of the Duchenne muscular
dystrophy gene in brain and other nonmuscle tissues
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 89 (12), 5346-5350 (1992)
MEDLINE 92302239
PUBMED 1319059
FEATURES Location/Qualifiers
source
1. .2110
/organism="Homo sapiens"
/db_xref="taxon:9606"
/map="Xp21.3-p21.1"
/cell_type="amniotic fluid"
/tissue_type="brain"
1. .2110
/gene="DMD"
1. .52
/gene="DMD"
/note="G00-119-850"
53. .1921
/gene="DMD"
/codon_start=1
/protein_id="AAA52316.1"
/db_xref="GI:181599"
/db_xref="GDB:G00-119-850"
/translation="MREQLKGHEQTTCWDHPKMTLYQSLADLNNVRFSAVRTAMKL
RRLQKALCLDLISAAACDALDQNLKQNDQPMDLIIINCLTIYDRLEQEHNNLVN
VPLCYDMCLNWLNVYDTGRTGRIRVLSFKTGIISLCAHLEDKRYLFFKQVASSTGF
CDQRRLLGLLHDSIQIPRLGEVASFGGSNIEPSVRSQFANNKPEIEAALFLDWMR
LEPQSMVWLPVLRVAAETAKHQAACNICKECPIIGFRYRSLKHPNYDQCSCFFSG
RVAKGKMHYPMVEYCTPTTSGEDVDFAKVLKKNKFKRYFAKHPRMGYLPVQTVLE
GDNMETPASSQLSHDDTHSRIEHYASRLAEMENSGSYLNDISPNFSDIDDEHLIIQ
HYCQSLNQDPLSQPRSPAQLISLESEERGERILADLEENRNLQAEYDRLLKQQH
EHKRLSPLSPPEMPTSPQSPRDAELIAEAKLLRQHKRLEARMQILEDHKNQLESQ
LHRLQLLEQPAEAKVNGTIVSSPSTLQSDSSQPMLLRVVGSQTSMSGEEDLLS
PPQDTSTGLEEVMEQLNNSPSPSRGHNVGSLFHMADDLGRAMESLSVSMTEEGAE"
1922. .2110
/gene="DMD"
/note="G00-119-850"
3'UTR
BASE COUNT 602 a 511 c 492 g 505 t
ORIGIN
Query Match 28.4%; Score 569; DB 9; Length 2110;
Best Local Similarity 100.0%; Pred. No. 1.8e-132;
Matches 569; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1433 CCACGAGACTCAAAACAACATTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTT 1492
Db 73 CCACGAGACTCAAAACAACATTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTT 132
QY 1493 AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT 1552
Db 133 AGCTGACCTGAATAATGTTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT 192
QY 1553 GCAGAAAGGCCCTTTTGGTGGATCTCTTGGCCCTGTGACGCTGATGCTGATGCTTGGACCA 1612

LOCUS AF339031 3521 bp mRNA linear VRT 20-APR-2001
DEFINITION Danio rerio dystrophin (dmd) mRNA, partial cds, alternatively spliced.
ACCESSION AF339031
VERSION AF339031.1 GI:13699249
KEYWORDS .
SOURCE Danio rerio.
ORGANISM Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes; Cyprinidae; Danio.
REFERENCE 1 (bases 1 to 3521)
AUTHORS Bolanos-Jimenez,F., Bordais,A., Behra,M., Strahle,U., Sahel,J. and Rendon,A.
TITLE Dystrophin and Dp71, two products of the DMD gene, show a different pattern of expression during embryonic development in zebrafish
JOURNAL Mech. Dev. 102 (1-2), 239-241 (2001)
MEDLINE 21184125
PUBMED 11287201
REFERENCE 2 (bases 1 to 3521)
AUTHORS Bolanos-Jimenez,F., Rendon,A. and Strahle,U.
TITLE Direct Submission
JOURNAL Submitted (22-JAN-2001) Laboratoire de Physiopathologie Retinienne, EMU 99-18 INSERM-Universite Louis Pasteur, 1, Place de l'Hopital, Strasbourg 67091, France
FEATURES
source
1. .3521
/organism="Danio rerio"
/db_xref="taxon:7955"
/chromosome="1"
/map="between z5508 and z5058"
<1. .3521
/gene="dmd"
<1. .3292
/gene="dmd"
/note="alternatively spliced"
/codon_start=2
/product="dystrophin"
/protein_id="AAK38376.1"
/db_xref="GI:13699250"
/translation="RPCRHWEDSHAKLTARVLTQLNMYKDSDDWLEARKRVEPLIKKANEKLESWKVSHSVEDLKQADVKQLSKDLQWQTMNVTNELANKLLTLYADDDTSKVKQMTESMNLAWANKKRGADKEADLEAGLROLOHYVLDLEKFLNWLTEAETANVLQDATRKEGLENPAIVRHLLQWQDLQAEIDAHRETHSLDENGHRIVSSLEGTDNAVVLQKRLDDMGQRWHELCNKVMSIRPYLDAGVDQWKLHLSLQELLNLWLKREELEKQKPVGGDPTVHQLLTHKAFRELGAKEPVINGTLDNAKTFLAEMPREGLKQRPQKQDVSPERVQNVGRILRKEVEDVTVRWKNLGAASVDWQOQLALERLMELODAQDQLDYKLQAESVKNKWPVGGELLVDLQNHIDRVKAFQEEIAPIQDNVNVNQLASTFRPSDLQSPDNLSRIDDLNMRWLLQISIEEHLSQLTAFKDLGSPQNFHASVESPLERSIQLSPNYYINHQTQTCWDHPKMAELYQSLADLNVRFSAYRTAMKLRMQKALCLDL LSPAAACEAFEQHNKQNEQFMDIVQVINCITSYDRLEQHSLSLVNPLCVDMLNWL LNNVYDTGRAGKIRTLSPKTIISLCKAHLDEKRYRFLFREVASATGFCQDQRRLLHLHDAIQPRQLGEVASFGSNIETPSVRSFCQFANNKPELEASVFLDWMRLPEQSMVWLPV LHRVAAAETAKHOAKCNICEKPIIGFRYRSLKHFNVDICQSCFFSGRVAKGHKMQYP MVEICTPTTSGEDVRDFAKVLNKNFKRTKRYFAKHPRMGYLPVQTILEGDNMETPVTLI NFWPVDHPPASSPOLSHDDTHSRIEHYASRDDEHLLIQHYCQSLNQGSPLSQPQSPAQ ILISMETEKEGELERVNLDLEQENRKLQAEYDRLLKKAHDHKLGLSPLPSPQMLPVSPQ SPRDAELIAEAKLLRQHKGRLEARMQILEDHKNQLESQRLRLQLLEQTESKVNQTAL SSPSTASPRSDTSLASLRVAASQTETMTGDDDELSSPTQDASTGLEVDIEQLNNSFPHS QGGRLNP"
3'UTR 3293. .3521
/gene="dmd"
BASE COUNT 1005 a 868 g 892 g 756 t
ORIGIN

Query Match 26.7%; Score 534; DB 5; Length 3521;
Best Local Similarity 71.4%; Pred. No. 1.2e-123;
Matches 718; Conservative 0; Mismatches 285; Indels 3; Gaps 1;
Qy 996 GCACAGACCCCTTGAAGACTCCAGGAACTTCAAGAGCCACGGATGAGCTGGACCTCAAG 1055
Db 1121 GAACTGGCCTTGGAGAGGCTGATGGAGCTCCAGGACGCCACGGATCAGCTGGACTACAAG 1180

QY 1056 CTGCGCAAGCTGAGGTGATCAAGGGATCTCTGGCAGCCCCGTGGGGGATCTCTCATTTGAC 1115
Db 1181 CTACGACAGGCTGAGAGCGTGAAGAAATCTTGGAAAGCCTGTTGGGAGCTGCTGTAGAC 1240
QY 1116 TCTCTCCAAGATCACCTCGAGAAAGTCAAGGCACTTCGAGGAGAAATTCGCCCTCTGAAA 1175
Db 1241 GATTTGCAAAACCATAGACAGAGTGAAGGCAATTCAGGAAGAGATAGCTCCCATTCAG 1300
QY 1176 GAGAACGTGAGCCACGTCATGACCTTGTCTGCCAGCTTACCACTTTGGGCATTCAGCTC 1235
Db 1301 GATAACGTAATCATGTTAACCGCTGGCCCTCCACATTTAGACCGCTCTGACATTCAGCTT 1360
QY 1236 TCACCGTATAACCTCAGCACTCTGGAAGACCTGAACACACAGATGGAAGCTTCTGCAGGTG 1295
Db 1361 TCTCCAGACACCTGAGCAGAAATCGATGACCTCAACATGAGATGGAGGCTTCTGCAGATC 1420
QY 1296 GCCGTGAGGACCGAGTCAGGCAGCTGCATGAAGCCACAGGGACTTTGGTCCAGCATCT 1355
Db 1421 TCCATTGAGGAGCATCTGAGTCAGCTGACCCACAGCCCTTAAGGACTTGGG---ACCCCTCT 1477
QY 1356 CAGCACTTTCTTTCCACGCTGTCTCCAGGGTCCCTGGGAGAGAGCCCATCTCGCCAAACAAA 1415
Db 1478 CAGAAATTTCTTTCATGCGCTGTGTTGAAAGCCCTCTCGAAGCCTCTATTTCAACCCAAACAT 1537
QY 1416 GTGCCCTACTATATCAACACGAGACTCAAAACAACTTGCTGGGACCATCCCAAAATGACA 1475
Db 1538 GTCCCTTACTATATCAATCAACACGAGCCCAACAAACATGTTGGGACCCACCCAAAGATGGCA 1597
QY 1476 GAGCTCTACCACTCTTTAGCTGACCTGAATAATGTGAGATCTCTGAGCCTTATAGGACTGCC 1535
Db 1598 GAACTCTACCACTGATAGCGGATCTCAACAAACGTCGCGGTTCTTCTGAGCATGCTGCAGCC 1657
QY 1536 ATGAAACTCCGAAGACTGCAGAAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTGAGCTGCA 1595
Db 1658 ATGAAGCTCAGACGAATGCAGAAAGGCCCTCTGTTTGGATCTTCTGAGCATGCTGCAGCC 1717
QY 1596 TGTGATGCCTTGGACCAGCACAACTCAAGCAAAATGACCAGCCCATGGATATCTCTGCAG 1655
Db 1718 TGTGAAGCCTTGGAGCAGCACAACTCAAAACAGAACGAGCAGTTCATGGACATCGTGCAG 1777
QY 1656 ATTATTAATTTGTTGACCACTATTATGACCGCCTGGAGCAAGAGCACAACTTTGGTCT 1715
Db 1778 GTATCAACTGTCTGACCAGCATCTACGACCGCTCTGGAGCAGCAGCAGCAGCCTGGTCT 1837
QY 1716 AACGTCCCTCTCTGCGTGGATATGTGTGAAGTGGCTGGTGAATGTTATGATACGGGA 1775
Db 1838 AACGTGCTCTCTGTGTGGACATGTGTCTCAACTGGCTGCTCAACGTTTACGATACAGGA 1897
QY 1776 CGAACAGGGAGGATCCGTGTCTCTTTAAAACTGGCATCATTTCCCTCTGTGTAAGCA 1835
Db 1898 CGAGCTGGGAAGATTCTGACCTATCTTCAAAACAGGAATAATCTCTTTGTGCAAAAGCT 1957
QY 1836 CATTGGAAGACAAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTGT 1895
Db 1958 CACCTTGAAGATAAGTACAGATTTTATTTTCGAGAGGTGGCCAGTGGCCACAGGCTTCTGT 2017
QY 1896 GACCAGCGCAGGCTGGCCCTCTCTTCTGATGATCTATCCAAATTTCCAAAGACAGTTGGGT 1955
Db 2018 GACCAGCGCGCCTCGGCCCTCTCTCTGATGATGCCATTCAGATCCCCAGGCGCTGGGT 2077
QY 1956 GAAGTTGCTATCTTTTGGGGGAGTAACTTGAAGCAAGTGTCCGGA 2001
Db 2078 GAAGTGGCGTCTTTTGGAGGGAGCAATATTGAGCCAGTGTGCGCA 2123

RESULT 12
HSDMDA1
LOCUS Human adult muscle mRNA fragment of DMD gene (DMD= Duchenne muscular dystrophy).
DEFINITION X06178 M22818
ACCESSION X06178.1 GI:30836
VERSION Duchenne muscular dystrophy.
KEYWORDS

Db 369 CGGATCTAGAAAAAGAAAAGCAATCCATCGGGCAAACACTGTATTCAATCAACAAGATCTTC 428

QY 892 TTCAACACTGAAGAATAAGTCAAGTGACCCAGAAAGACGGAACCATGGCTGGATAA CTTTG 951
|||||

Db 429 TTCAACACTGAAGAATAAGTCAAGTGACCCAGAAAGACGGAACCATGGCTGGATAA CTTTG 488
|||||

QY 952 CCCGGTGTGGGATAATTTAGTCCAAAAA ACTTGAAAAGACTACAGCACAGACCCTT 1007
|||||

Db 489 CCCGGTGTGGGATAATTTAGTCCAAAAA ACTTGAAAAGACTACAGCACAGATGATT 544
|||||

RESULT 13
RNAPDY3
LOCUS RNAPDY3 1240 bp mRNA linear ROD 11-AUG-1993
DEFINITION R.norvegicus mRNA for APO-dystrophin-3.
ACCESSION X69767
VERSION X69767.1 GI:57913
KEYWORDS apodystrophin.
SOURCE Rattus norvegicus.
ORGANISM Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.

REFERENCE 1 (bases 1 to 1240)
AUTHORS Tinsley,J.M.
TITLE Direct Submission
JOURNAL Submitted (14-DEC-1992) J.M. Tinsley, Molecular Genetics Group,
Institute of Molecular Medicine, John Radcliffe Hospital,
Headington, Oxford OX3 9DU, UK

REFERENCE 2 (bases 1 to 1240)
AUTHORS Tinsley,J.M., Blake,D.J. and Davies,K.E.
TITLE Apo-dystrophin-3: a 2.2kb transcript from the DMD locus encoding
the dystrophin glycoprotein binding site
JOURNAL Hum. Mol. Genet. 2 (5), 521-524 (1993)
MEDLINE 93299455
PUBMED 8518789

FEATURES source Location/Qualifiers
1..1240
/organism="Rattus norvegicus"
/db_xref="taxon:10116"
/cell_line="Schwannowa cells"
/dev_stage="neonatal"
106..1128
/codon_start=1
/product="apodystriophin-3"
/protein_id="CAA49423.1"
/db_xref="GI:57914"
/db_xref="SPTREMBL:Q05485"
/translation="MREHLKGHETQTTCWDHPKMTELYQS LADLN NVRFSAYRTAMKL
RRLQKALCLDLLSLSA CDALDQHNLKNDQPMDILOI IYCLTIYDRLEQEHN NLVN
VPLCVDCVLNLLNVYDTGRTGRIRVLSFKTGII SLCKAHLEDKYRYLFKQVASSTGF
CDORRLGLLHDISI QIPQLGEVASFGGSNI EPSVRS CFQFANNKP EIEA ALFLDMNR
LE PQSMWL PVLHRVA AEAKHQA KCNICKECPIIGFRYSRLKHFN YDICQSCFFSG
RVAKGHKMHPMVVEYCTPTTSGEDVRDFAKVLKNKFTRKRYFAKHPRMGYLPVQTIVLE
GDNMET"

CDS BASE COUNT 323 a 314 c 287 g 316 t
ORIGIN

Query Match 25.9%; Score 518; DB 10; Length 1240;
Best Local Similarity 90.9%; Pred. No. 1.3e-119;
Matches 551; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 1396 GAGCCATCTGCCCAAACAAAGTGCCTTACTATATCAACCACGAGACTCAAACAACTTGCT 1455
| | ||||| ||| || | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 89 GTGAATCCTTACAACCATGAGGGAACACCTCAAAGGCCACGAGACCCAAACAACTTGTT 148

QY 1456 GGGACCATCCCCAAAATGACAGAGTCTTACCAGTCTTTTAGCTGACCTGAATAATGTCAGAT 1515
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 149 GGGACCATCCCCAAAATGACAGAGTCTTACCAGTCTTTAGCTGACCTGAATAATGTCAGGT 208

QY 1516 TCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAGGCCCTTTGCTTGGATC 1575
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 209 TCTCCGCTTATAGGACTGCCATGAAGCTCAGAAGACTCCAGAAGGCCCTTTGCTTGGATC 268

QY	1576	TCCTGAGCCCTGCAGCTGCATGTGATGCCCTTGACCAGCACAAACCTCAAGCAAAATGACC	1635
Db	269	TCCTGAGCCCTGCAGCTGCATGTGACGCCCTTGACCAGCACAAACCTCAAGCAAAATGATC	328
QY	1636	AGCCCATGGATATCCTGCAGATTATTAAATGTTTGACCACCTATTATGACCCGCTGGAGC	1695
Db	329	AGCCCATGGATATCCTGCAGATAATTACTGTTTGACTACTATTATGACCCGCTGGAGC	388
QY	1696	AAGAGCACACAATTTGGTCAACGTCCTCTCTGCGTGGATATGTGTCTGAACCTGGCTGC	1755
Db	389	AAGAGCACACAATTTGGTCAATGTCCCTCTCTGTGTGGATGTGTGTCTCAACTGGCTTC	448
QY	1756	TGAATGTTTATGATACGGGACGAACAGGGAGGATCCGCTGCTCTCTTTTAAACTGGCA	1815
Db	449	TCAATGTTTATGATACGGGACGAACAGGGAGGATCCGCTGCTCTCTTTTAAACTGGCA	508
QY	1816	TCATTTCCCTGTGTAAGCACACATTTGGAAGACAAGTACAGATACCTTTTCAAGCAAGTGG	1875
Db	509	TCATTTCTCTGTGTAAGCACACACTTGAAGACAAGTACAGATACCTTTTCAAGCAAGTGG	568
QY	1876	CAAGTTCAACAGGATTTGTGACCAGCGCAGGCTGGGCTCCTTCTGTCATGATTTCTATCC	1935
Db	569	CAAGTTCAACCCGGCTTTGTGACCAGCGTAGACTGGGTCTTCTTCTGTCATGATTTCTATTC	628
QY	1936	AAATTTCCAAAGACAGTTGGTGAAAGTTGCATCCTTTGGGGCGCAGTAACATTTGAGCCAAAGTG	1995
Db	629	AAATCCCAAGACAGTTGGTGAAAGTTGCCTCCTTTGGGGCGCAGTAACATTTGAGCCGAGTG	688
QY	1996	TCCGGA 2001	
Db	689	TCAGGA 694	
RESULT 14			
MMGUTRPH			
LOCUS			
DEFINITION			
ACCESSION			
VERSION			
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REMARK			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
FEATURES			
source			
1. .3161			
/organism="Mus musculus"			
/db_xref="taxon:10090"			
/clone="AU#3"			
/tissue_type="brain"			
/clone_lib="adult mouse brain"			
/dev_stage="adult"			
1. .197			
198. .3161			
/function="Unknown"			
/standard_name="G-utrophin"			
/citation=[1]			
5'UTR			
CDS			

/codon_start=1			
/evidence=experimental			
/product="G-utrophin (predicted protein)"			
/protein_id="CAA58496.1"			
/db_xref="GI:793852"			
/db_xref="MGD:MG1:104631"			
/db_xref="SPTREMBL:Q61636"			
/translation="MQILRLTQKCGKLMMAVVRTSLQKVVVLLHRLQRMNAVSSPRVQ			
KLCKDIOAEIDAHNDFKSIDGNRQKMYKALGNSEETMLQHRLLDDMNQRWNLKAKS			
ASIRAHLEASAEKWNRRLLASLEELIKWLNKMKDEELKKQMPIGGDVPALQLOYDHCKVL			
RRELKEKEYSVLNAVDAQVFLADQPIEAPPEPRNPQSKTELTPEERAQKIAKAMRK			
QSSEVREKWEENLNAVTSNWQKQVKALEKLRDLQGGAMDDLDADMKVEAVRNWKPVG			
DLLLDSLDQHIEKTLAFREEIAPINLKVKTMDLSSQLSPLDLHPSLUKMSRQLDLDLNM			
RWKLLQVSVDDRLKQLQEAHRDFGPPSQHFLSTSVQLPWQRSISHNKVPYYINHQTQT			
TCWDHPKMTLFSQSLADLNNVRFSAYRTAKIRRLQKALCLDLLELNTTNEVFKQHKLL			
NQNDQLLSVPDVINCLTITYDGLEQLHKDLVNVPLCVDMCLNWLNNVYDTRTGKIRV			
QSLKIGLSLSKGLLEEKYRCLFEKVEAGPTMCDQRLGLLLHDAIQIPROLGEVAAF			
GGSNIETPSVRSFQNNNKPEISVKEFIDWMHLEPQSMVWLPVLHRAAAETAKHQAK			
CNICKECPVGFPRYRSLKHFNYDVCQSCFFSGRTAKGHKLHYPMVEYCIPTTSGEDVR			
DFTKVLKNFRSKKYFAKHPRLGYLPVQTVLEGNLETPITLISMWPEHYDPSQSQL			
FHDDTHSRIEQYATRLAQMERTNGSFLTDSSTTGSVEDEHALIQYQCTLGGESPV			
FQPSPAQILKSVEREEGELERIADLEEEQRNLQVEYEQLKEQHLRRLGLPVGSPD			
IVSPHHTSEDSIELIAEAKLLRQHKRLEARMQILEDHNKQLESQHLRLQLLEQDSD			
SRINGVSPWASPOHSALSYSLDTPGPQPHQAASEDLLAPPHDTSTDLTQDMEQIN			
FPSCSNVPSRPQAM"			
BASE COUNT	887 a	784 c	802 g 688 t
ORIGIN			
Query Match	25.3%;	Score 506.6;	DB 10; Length 3161;
Best Local Similarity	68.6%;	Pred. No. 9.8e-117;	
Matches 698;	Conservative 0;	Mismatches 319;	Indels 0; Gaps 0;
Qy	985	AAAAGAGTACAGCACAGACCCCTTGAAGAGACTCCAGGAACCTTCAAGAGGCCACGGATGAGC	1044
Db	910	AAAAGCAAGTAGGGAAGCGGTTAGAGAAACTCCGAGACCTGCAGGGAGCTATGGACGACC	969
Qy	1045	TGGACCTCAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGGCGATC	1104
Db	970	TGGACGCAGACATGAAGGAGGTGGAGGCTGTGCGGAATGGCTGGAAGCCCGTGGGAGACC	1029
Qy	1105	TCCTCATTTGACTCTCTCCAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAATTG	1164
Db	1030	TGCTTATAGACTCCCTGCAGGATCACATCGAGAAACCCCTGGCGTTTAGAGAGAAATTG	1089
Qy	1165	CGCCTCTGAAAGAGAACGTGAGCCACGTCAATGACCTTGCTCGCCAGCTTACCACCTTGG	1224
Db	1090	CACCAATCAACTTAAAGTAAACAAATGAATGACCTGTCCAGTCAGCTGTCTCCACTTG	1149
Qy	1225	GCATTGACGCTCTCACCGTATAACCTTCAGCAGCTCTGGAAGACCTGAACACCCAGATGGAAGC	1284
Db	1150	ACTTGCATCCATCTCTAAAGATGTCTCGCCAGCTGGATGACCTTAATATGCGATGGAAC	1209
Qy	1285	TTCTGCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGAAGCCCAACAGGACTTTG	1344
Db	1210	TTCTACAGGTTTCCGTGGACGATCGCCTTAAGCAGCTCCAGGAAGCCCAACAGAGATTTTG	1269
Qy	1345	GTCCAGCATCTCAGCAGCTTTCTTTCACAGTCTGTCTCCAGGTCCTCGGAGAGAGCCATCT	1404
Db	1270	GGCCATCTTCTCAACACTTTCTGTCCACTTCAGTCCAGCTGCCGTGGCAGAGATCCATTT	1329
Qy	1405	CGCCAAACAAAGTGCCCTACTATATCAACACAGAGACTCAAAACAACTTGTGGGACCATC	1464
Db	1330	CACATAATAAGTGCCCTATTACATCAACCATCAACACACAGACAACCTGTTGGGATCATC	1389
Qy	1465	CCAAATGACAGAGCTCTACCAGTCTTTTAGCTGACCTGAATAATGTACAGATTCTCAGCTT	1524
Db	1390	CTAAATGACTGAGCTCTTCCCAATCCCTTGCTGATCTGAATAATGTACGTTTCTCTGCCT	1449
Qy	1525	ATAGGACTGCCATGAAACTCCGAAGACTGCAGAAGGCCCTTTGCTTGGATCTCTTGAGCC	1584
Db	1450	ACCGCACAGCAATCAAAATTCGAAGGCTGCAAAAAGCATTATGTCTGGATCTCTTAGAGC	1509
Qy	1585	TGTCAGCTGCATGTGATGCCTTGGACCAGCACAACTCAAGCAAAATGACCAGCCCATGG	1644

1510 TGAATACGACGAATGAAGTTTTCAGCGACGACACAACTGAACCAAAATGATCAGCTCCTGA 1569
1645 ATATCCTGCAGATTATTAATTTGTTGACCACTATTATGACCGCCTGGAGCAAGACACA 1704
1570 GTGTCCCAGACGTCATCAACTGTCTGACCACTTACGATGGCTTGAGCAGCTGCACA 1629
1705 ACAATTTGGTCAACGTCCTCTCTGCGTGGATATGTGTCTGAACCTGGCTGCTGAATGTTT 1764
1630 AGGACTTGGTCAATGTTCCACTCTGCGTCGATATGTGTCTCAACTGGCTGCTCAACGTA 1689
1765 ATGATACGGGACGACAGAGGAGGATCCGTGTCTCTCTTTTAAAACTGGCATCATTTCCC 1824
1690 ACGACACGGGCGGACTGGAAAAATTCGGGTACAGAGTCTGAAGATTGGATTGATGTC 1749
1825 TGTGTAAGCACATTTGGAAGACAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAA 1884
1750 TCTCCAAAGGCCCTCTTAGAAGAGAAATACAGATGTCTCTTTAAAGGAGGTGGCAGGGCCAA 1809
1885 CAGGATTTTGTGACACGCGCAGGCTGGCCTCTCTCTGTCATGATCTATCCAAATCCAA 1944
1810 CTGAGATGTGTGACACGCGCAGCTTGGCCTGCTACTTACGATGCCATCCAGATCCCTA 1869
1945 GACAGTTGGGTGAAGTTGCATCCTTTGGGGGACGATAACATGAGCCAAAGTGTCCGGA 2001
1870 GGCAGCTGGGGGAAGTAGCAGCCCTTTGGGGGACATAACATTGAGCCCAGTGTCCGCA 1926
RESULT 15
AF195788
LOCUS
DEFINITION
Rattus norvegicus dystrophin-related protein 2 B-form splice
variant (Drp2) mRNA, complete cds.
AF195788
AF195788.1 GI:11066166
Rattus norvegicus.
Rattus norvegicus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
Rattus.
1 (bases 1 to 3073)
Roberts,R.G. and Sheng,M.
Association of dystrophin-related protein 2 (DRP2) with
postsynaptic densities in rat brain
Mol. Cell. Neurosci. 16 (5), 674-685 (2000)
20538711
11083927
2 (bases 1 to 3073)
Roberts,R.G. and Sheng,M.
Direct Submission
Submitted (18-OCT-1999) Division of Medical and Molecular Genetics,
GKT Medical School, 8th Floor, Guy's Hospital, Guy's Tower, London
SE1 9RT, UK
Location/Qualifiers
1. .3073
/organism="Rattus norvegicus"
/strain="Sprague-Dawley"
/db_xref="taxon:10116"
1. .3073
/gene="Drp2"
69..70
/gene="Drp2"
/note="splice acceptor AG dinucleotide used by A-form"
200..3073
/gene="Drp2"
/note="DRP2; membrane-associated cytoskeletal protein;
contains two spectrin repeats; WW domain; Z2 domain"
/codon_start=1
/product="dystrophin-related protein 2 B-form splice
variant"
/protein_id="AAG28485.1"
/db_xref="GI:11066167"

/translation="MQPLVMQGPCYTLPRCHEWHAADREHHSSSLRNTCPQPVRAAV
TIPAPWDGADGPCLSPKLLNGSVGAVPLEPSAMNLCWNEIKKSHNLRARLEAFSD
HSGKLQPLQEIIDWLSQKDELSAQLPLOGDVALVQEKETHAAFMEEVKSGPYIY
SVLESAQAFLSQHPFEELESSES KDTSPRIQNLSRFVWKOATVASELWEKLTAR
CVDQHRHIEHTLEHLLIEQAMEELSSTLTQAEGRATWEPIGDLFIDSLPEHIOAIK
LFKEEFPVKDGVKLVNDLAHQLAISDVHLSMENSRALEQINVRWKOLQVSVARLQK
LQDAHRDFGPGSQHFSTSVQVPERAISPKNVPYIYINHQATTCWDHPKMTELYQTL
ADLNNIKFSAYRTAMKLRVQKALRIDLVLTLTALEIFNEHDLQASEHVMVDVVEIHC
LTALYERLEERGILVNVPLCDMSLNWLLNVDSPRGSKMRALSFKTIACLCGTEV
KEKQLYLFQVANSKGDORHLGALLHEATQVPRQLGSEVAAFGSNVPSVRSCEF
STGKPVIEASQFLEWNLBPQSMVWLAVLHRTVAEQVKHQTCSICRQCPIKGFYR
SLKQFNVDICQTCFLTGRASKGNKLHYPIMEYYTPTTSENMRDFAATLKNKFRSKQY
FSKHPQRYLPVQSVLESDCSETPASSPMLPHADTHSRIEHFAASRLAEMESQNCSEFN
DSLSPDDSIDEDQYLLRHSSPTIDREPAFQQAQPCSMATESKGELEKILAHLEDENRI
LOGELRRLKWQHEEAVEAPTAEASAEATPDHRNEELLAEARILRQHKSRLETRMOIL
EDHNKQLESQLORLRELLLOPPTESDNGSAGSSSLASSPRQSEGSHPREKQTTPTDTE
AADDVGSKSQDVSHSLEDIMEKLRHAFPSVRSSDVTANTLLAS"
BASE COUNT 779 a 843 c 790 g 659 t 2 others
ORIGIN
Query Match 23.7%; Score 475; DB 10; Length 3073;
Best Local Similarity 67.1%; Pred. No. 9.1e-109;
Matches 673; Conservative 0; Mismatches 330; Indels 0; Gaps 0;
QY 999 CAGACCCCTTGAAAGACTCCAGGAACCTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTG 1058
Db 881 CACACCCCTGGAACATCTATTGGAGATCCAAAGGGCAATGGAGAACTGAGCAGTACTTTG 940
QY 1059 CGCCAAAGCTGAGGTGATCAAGGGATCTCTGGCAGCCCGTGGCGGATCTCCTCATTTGACTCT 1118
Db 941 ACCCAAGCAGAGGGAGTCCGAGCCACATGGAGCCCATAGGAGATCTCTTTATCGATTCC 1000
QY 1119 CTCCAAGATCACCTCGAGAAAGTCAAGGCACTTCGAGGAGAGAAATTCGCCTCTGAAAGAG 1178
Db 1001 CTCCCAGAGCATATCCAAAGCCATCAAGTTATTCAAAGAAAGAAATTTCTCTCTGTGAAAGAC 1060
QY 1179 AACGTGAGCCACGTCATGACCTTGCTGCGCCAGCTTACCACCTTTGGGCAATTCAGCTCTCA 1238
Db 1061 GGGGTGAAGTTAGTGAATGACCTGGCCACACGCTTGTATTTCTGATGTGCACTTGTC 1120
QY 1239 CCGTATAACCTCAGCACTCTGGAAGACCTGAACACACAGATGGAAGCTTCTGCAGGTGGCC 1298
Db 1121 ATGGAGAATTCAAGGGCTCTGGAACAGATCAACGCTCCGCTGGAACACAGCTCCAGGTGTC 1180
QY 1299 GTCGAGGACCGAGTCAAGGCACTGCAATGAAGCCCAAGGCACTTTGGTCCAGCATCTCAG 1358
Db 1181 GTTGTGAGAGGGCTTAAACAGCTCCAGGATGCCACCGGCACTTTGGGCGCTGGGTACAG 1240
QY 1359 CACTTTCTTTCCACGCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTG 1418
Db 1241 CACTTCTCTCCACTTCTGTTCAGTTCCCTGGGAAAGAGCAATTTCTCCCAATAAAGTC 1300
QY 1419 CCCTACTATATCAACACGAGACTCAACAACTTGTGAGCCATCCCAAAATGACAGAG 1478
Db 1301 CCCTACTACATCAACACAGGCTCAGACCCACATGCTGGGACCATCCCAAGATGACCGAG 1360
QY 1479 CTCTACAGTCTTTAGCTGACCTGAATAATGTGAGATTTCTCAGCTTATAGGACTGCCATG 1538
Db 1361 TTATACCAAAACCTAGCTGACCTGAACAACTCAAGTTCTCAGCTTACCGCACTGCCATG 1420
QY 1539 AAACCTCCGAAGACTGCAGAAAGGCCCTTTGCTTGGATCTCTTGAGCCCTGTGAGCTGCATG 1598
Db 1421 AAGCTCCGCGAGTCCAGAGGCCCTCGCTTGGATCTGGTAACCTTAACCTACAGCTCTG 1480
QY 1599 GATGCTTGGACCAAGCAACCTCAAGCAAAATGACCAGCCCATGGATATCTCTGCAGATT 1658
Db 1481 GAGATCTTCAATGACACGACTTGCAGGCCAGTGAACATGTAATGGATGTGGTGGAGGTC 1540
QY 1659 ATTAATTGTTGACCACTATTATGACCGCCTGGAGCAAGAGCACAAATTTGGTCAAC 1718
Db 1541 ATTCACTGCTTGAAGTGCCTTGTATGAACGACTGGAGGAGGAAAGAGGATCTCTGGTCAAC 1600
QY 1719 GTCCCTCTCTGCGTGGATATGTGTGAACTGGCTGCTGAATGTTTATGATACGGGACGA 1778

Db 1601 GTGCCGCTGTGTAGACATGAGCCTCAACTGGGTCCTCAACGTTTTTGATAGTGGCCGC 1660
QY 1779 ACAGGAGGATCCCGTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACAT 1838
Db 1661 AGTGGAAAGATGCCGAGCATTTGTCCTTTAAGACTGGCATCGCATGCCTGTGTGGCACATGAA 1720
QY 1839 TTGGAAGACAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTIGIGAC 1898
Db 1721 GTGAAGAAAAAACTTCAGTATCTCTTCAGCCCAAGTGGCCCAATTTCAGGCAGCAAGTGTGAC 1780
QY 1899 CAGCGCAGGCTGGGCCCTCCTTCTGCATGATTCTATCCAAAATTCCAAGACAGTTGGGTGAA 1958
Db 1781 CAGCGCCACCTCGGTGCCCTGCTTCATGAAGCCATCCAAGTGCCCCCGTCAGCTGSGGTGAA 1840
QY 1959 GTTGCATCCTTTGGGGGCGAGTAACATTGAGCCCAAGTGTCCCGA 2001
Db 1841 GTGGCAGCATTTGGGGGCGAGCAATGTGGAGCCCCAGTGTCCGTA 1883

Search completed: April 25, 2003, 14:29:27
Job time : 3446.67 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 07:25:57 ; Search time 279.019 Seconds
(without alignments)
16150.348 Million cell updates/sec

Title: US-09-845-416-14_COPY_1000_3000
Perfect score: 2001
Sequence: 1 gccagttcattgatggagag.....cattgagccaagtgtccgga 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4313688

Minimum DB seq length: 0
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_101002:*
1: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDS2/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2001	100.0	3446	24	AAD37242
2	1990	99.5	4414	24	AAD37260
3	1593	79.6	3510	24	AAD37240
4	1593	79.6	4476	24	AAD37259
5	1332	66.6	4402	21	AAZ48568
6	1269.2	63.4	4402	21	AAZ48567
7	1263	63.1	3858	24	AAD37237
8	1263	63.1	4825	24	AAD37257
9	1263	63.1	4848	24	AAD37263

10	1251.2	62.5	3531	24	AAD37238	Human dystrophin m
11	1251.2	62.5	4498	24	AAD37258	Adeno-associated v
12	1245	62.2	4182	24	AAD37230	Human dystrophin m
13	1112	55.6	3999	24	AAD37234	Human dystrophin m
14	1112	55.6	4966	24	AAD37256	Adeno-associated v
15	1112	55.6	4990	24	AAD37262	Adeno-associated v
16	1002.6	50.1	1821	24	AAD37241	Human dystrophin r
17	1002.6	50.1	2169	24	AAD37232	Human dystrophin r
18	1001	50.0	1434	24	AAD37243	Human dystrophin r
19	985.8	49.3	1991	24	AAD37231	Human dystrophin N
20	767	38.3	3275	10	AAN97129	Partial sequence o
21	735.4	36.8	4075	21	AAZ48569	A rod shortened dy
22	731.2	36.5	3747	21	AAZ48566	A rod shortened dy
23	677.6	33.9	3163	21	AAZ48571	A rod shortened dy
24	665	33.2	1667	24	AAD37235	Human dystrophin N
25	458.8	22.9	3172	21	AAZ48570	A rod shortened dy
26	457	22.8	887	24	ABK81993	Human dystrophin C
27	350	17.5	1340	24	AAD37239	Human dystrophin N
28	327	16.3	327	24	ABK81966	Human dystrophin s
29	324	16.2	324	24	ABK81991	Human dystrophin s
30	322	16.1	333	24	ABK81968	Human dystrophin s
31	321	16.0	333	24	ABK81967	Human dystrophin s
32	215	10.7	215	24	ABK81992	Human dystrophin h
33	186.2	9.3	1500	22	AAF84672	Nucleotide sequenc
34	180	9.0	476	24	ABS08442	Human genome-deriv
35	160.2	8.0	256	22	ABA69736	Human foetal liver
36	160.2	8.0	256	22	ABA36636	Probe #15102 for g
37	160.2	8.0	256	22	AAK17925	Human brain expres
38	160.2	8.0	256	22	AAK43799	Human bone marrow
39	160.2	8.0	256	22	AAI24574	Probe #14507 for g
40	160.2	8.0	256	22	AAI49820	Probe #18506 used
41	160.2	8.0	256	24	ABSL8027	Human genome-deriv
42	159.2	8.0	4928	23	ABL12611	Drosophila melanog
43	144	7.2	466	22	ABA58628	Human foetal liver
44	144	7.2	466	22	ABA27626	Probe #6092 for ge
45	144	7.2	466	22	AAK06756	Human brain expres

ALIGNMENTS

RESULT 1
AAD37242
ID AAD37242 standard; DNA; 3446 BP.
XX
AC AAD37242;
XX
DT 21-AUG-2002 (first entry)
XX
DE Human dystrophin minigene delta3447.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Homo sapiens.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
XX
PI Xiao X;
XX
DR WPI; 2002-049342/06.
XX
PT New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,

rod repeats, H1 and H4 domains and a cysteine rich domain of a dystrophin gene

Example 1; Page 53-54; 7lpp; English.

The present invention relates to an isolated nucleotide sequence encoding a dystrophin minigene. The minigene comprises N-terminal or modified N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4 domains and cysteine-rich domains of dystrophin or utrophin genes. The invention also relates to a recombinant adeno-associated virus (AAV) comprising dystrophin minigene operably linked to an expression control element. The dystrophin minigene in operable linkage with an expression control element, in a recombinant adeno-associated virus or retrovirus is useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is human dystrophin minigene delta3447 containing nucleotides 1-1992 (N-terminus, hinge H1 and rods R1, R2 and R3), 8749-10227 (rod R24, hinge H4 and CR domain) and 11047-11058 (dystrophin last 3 amino acids).

Sequence 3446 BP; 1074 A; 766 C; 787 G; 819 T; 0 other;

Query Match 100.0%; Score 2001; DB 24; Length 3446;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 2001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	GGCAGTTCATTGATGGAGAGTGAAGTAAACCTGGACCGTTATCAACACAGCTTTAGAAGAA	60
Db	1000	GGCAGTTCATTGATGGAGAGTGAAGTAAACCTGGACCGTTATCAACACAGCTTTAGAAGAA	1059
QY	61	GTATTATCGTGGCTTCTTCTGCTGAGGACACATTGCAAGCACACAGGAGAGATTCTTAAT	120
Db	1060	GTATTATCGTGGCTTCTTCTGCTGAGGACACATTGCAAGCACACAGGAGAGATTCTTAAT	1119
QY	121	GATGTGGAAGTGGTGAAGACCAAGTTCATCTACTGAGGGGTACATGATGGATTGACA	180
Db	1120	GATGTGGAAGTGGTGAAGACCAAGTTCATCTACTGAGGGGTACATGATGGATTGACA	1179
QY	181	GCCCATCAGGGCCGGGTTGGTAATATCTACAAATGGGAAGTAAGCTGATTGGAACAGGA	240
Db	1180	GCCCATCAGGGCCGGGTTGGTAATATCTACAAATGGGAAGTAAGCTGATTGGAACAGGA	1239
QY	241	AAATTATCAGAAGATGAAGAACTGAAGTACAAAGACAGATGAATCTCTAAATTCAGA	300
Db	1240	AAATTATCAGAAGATGAAGAACTGAAGTACAAAGACAGATGAATCTCTAAATTCAGA	1299
QY	301	TGGGAATGCCTCAGGGTAGCTAGCATGGAACAAACAAAGCAATTTACATAGAGTTTAATG	360
Db	1300	TGGGAATGCCTCAGGGTAGCTAGCATGGAACAAACAAAGCAATTTACATAGAGTTTAATG	1359
QY	361	GATCTCCAGAAATCGAAACTGAAAGAGTTGAATGACTGGCTAAACAAAACAGAAAGAA	420
Db	1360	GATCTCCAGAAATCGAAACTGAAAGAGTTGAATGACTGGCTAAACAAAACAGAAAGAA	1419
QY	421	CAAGGAAATGGAGGAAGAGCCCTCTTGGACCTGATCTTGAAGACCTAAACCGCCAAAGTAC	480
Db	1420	CAAGGAAATGGAGGAAGAGCCCTCTTGGACCTGATCTTGAAGACCTAAACCGCCAAAGTAC	1479
QY	481	AACAACATAAGGTGCTTCAAGAAGATCTAGAACAAAGCAAGTCAGGGTCAATCTCTCA	540
Db	1480	AACAACATAAGGTGCTTCAAGAAGATCTAGAACAAAGCAAGTCAGGGTCAATCTCTCA	1539
QY	541	CTCACATGGTGGTGGTAGTTGATGAATCTAGTGGAGATCAGGCAACTGCTGCTTTGGAAG	600
Db	1540	CTCACATGGTGGTGGTAGTTGATGAATCTAGTGGAGATCAGGCAACTGCTGCTTTGGAAG	1599
QY	601	AACAACCTTAAGGTATTTGGGAGATCGATGGGCAACATCTGTAGATGGACAGAAGACCGCT	660
Db	1600	AACAACCTTAAGGTATTTGGGAGATCGATGGGCAACATCTGTAGATGGACAGAAGACCGCT	1659
QY	661	GGGTTCTTTTACAAGACATCCTTCTCAAAATGGCAACGCTTACTGAAGAACAGTGCCTTT	720
Db	1660	GGGTTCTTTTACAAGACATCCTTCTCAAAATGGCAACGCTTACTGAAGAACAGTGCCTTT	1719

QY	721	TTAGTGCATGGCTTTTCAGAAAAAGAGATGACAGTGAACAAGATTACACAACTGGCTTTA	780
Db	1720	TTAGTGCATGGCTTTTCAGAAAAAGAGATGACAGTGAACAAGATTACACAACTGGCTTTA	1779
QY	781	AAGATCAAAATGAAATGTTATCAAGTCTTCAAAAACTGGCCGTTTTAAAAAGCGGATCTAG	840
Db	1780	AAGATCAAAATGAAATGTTATCAAGTCTTCAAAAACTGGCCGTTTTAAAAAGCGGATCTAG	1839
QY	841	AAAAGAAAAGCAATCCATGGGCAAACTGTATTCACTCAAAACAAGATCTTCTTTCAACAC	900
Db	1840	AAAAGAAAAGCAATCCATGGGCAAACTGTATTCACTCAAAACAAGATCTTCTTTCAACAC	1899
QY	901	TGAAGAATAAGTCAGTACCCAGACGGAAGCATGGCTGGATAACTTTGCCCGGTGT	960
Db	1900	TGAAGAATAAGTCAGTACCCAGACGGAAGCATGGCTGGATAACTTTGCCCGGTGT	1959
QY	961	GGGATAATTTAGTCCAAAAAATTTGAAAAGAGTACAGCACAGACCCCTTGAAAAGACTCCAGG	1020
Db	1960	GGGATAATTTAGTCCAAAAAATTTGAAAAGAGTACAGCACAGACCCCTTGAAAAGACTCCAGG	2019
QY	1021	AACTTCAAGAGGCCACGATGAGCTGGACCTCAAGCTGGCCCAAGCTGAGGTGATCAAGG	1080
Db	2020	AACTTCAAGAGGCCACGATGAGCTGGACCTCAAGCTGGCCCAAGCTGAGGTGATCAAGG	2079
QY	1081	GATCCTGGCAGCCCGTGGCGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGAGAAAG	1140
Db	2080	GATCCTGGCAGCCCGTGGCGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGAGAAAG	2139
QY	1141	TCAAGGCACCTTCGAGGAGAAATTCGCCCTCTGAAAAGAGAACGTGAGCCACGTCAATGACC	1200
Db	2140	TCAAGGCACCTTCGAGGAGAAATTCGCCCTCTGAAAAGAGAACGTGAGCCACGTCAATGACC	2199
QY	1201	TTGCTCGCCAGCTTACCACCTTTGGGCATTCAGCTCTCACCGTATAACCTCAGCACTCTGG	1260
Db	2200	TTGCTCGCCAGCTTACCACCTTTGGGCATTCAGCTCTCACCGTATAACCTCAGCACTCTGG	2259
QY	1261	AAGACCTGAACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGGCAGC	1320
Db	2260	AAGACCTGAACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGGCAGC	2319
QY	1321	TGCATGAAGCCCAACAGAGGACCTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGTCTGTCC	1380
Db	2320	TGCATGAAGCCCAACAGAGGACCTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGTCTGTCC	2379
QY	1381	AGGTCCCTGGGAGAGAGCCCATCTCGCCAAACAAAGTGCCCTACTATATCAACACGAGAG	1440
Db	2380	AGGTCCCTGGGAGAGAGCCCATCTCGCCAAACAAAGTGCCCTACTATATCAACACGAGAG	2439
QY	1441	CTCAACAACTTGTCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACC	1500
Db	2440	CTCAACAACTTGTCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACC	2499
QY	1501	TGAATAATGTCAGATTCCTCAGCTTATAGGACTGCCATGAAACTCCGAAAGACTGCAGAAGG	1560
Db	2500	TGAATAATGTCAGATTCCTCAGCTTATAGGACTGCCATGAAACTCCGAAAGACTGCAGAAGG	2559
QY	1561	CCCTTTGCTTGGATCTCTTGAGCCCTGTGAGTGCATGTGATGCCCTTGGACCAGCACAAACC	1620
Db	2560	CCCTTTGCTTGGATCTCTTGAGCCCTGTGAGTGCATGTGATGCCCTTGGACCAGCACAAACC	2619
QY	1621	TCAAGCAAAATGACAGCCCATGGATATCCTGCAGATTAATTTGTTGACCACTATTT	1680
Db	2620	TCAAGCAAAATGACAGCCCATGGATATCCTGCAGATTAATTTGTTGACCACTATTT	2679
QY	1681	ATGACCGCCTGGAGCAAGAGACACAACAATTTGGTCAACGTCCTCTCTGCGTGGATATGT	1740
Db	2680	ATGACCGCCTGGAGCAAGAGACACAACAATTTGGTCAACGTCCTCTCTGCGTGGATATGT	2739
QY	1741	GTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGACAGAGGAGGATCCGTGTCTGT	1800
Db	2740	GTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGACAGAGGAGGATCCGTGTCTGT	2799
QY	1801	CTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAGATACC	1860

Db 1360 GATCTCAGAAATCAGAAACTGAAAGAGTTGATGACTGGCTAACAAAACAGAGAAAGA 1419
QY 420 ACAAGGAAAATGGAGGAAGAGCCCTCTTGGACCTGATCTTTGAAGACCTAAAACGCCAAGTA 479
Db 1420 ACAAGGAAAATGGAGGAAGAGCCCTCTTGGACCTGATCTTTGAAGACCTAAAACGCCAAGTA 1479
QY 480 CAACAACATAAAGGTGCTTCAAGAAGATCTAGAACAAGAACAAAGTCAGGGTCAATTCTCTC 539
Db 1480 CAACAACATAAAGGTGCTTCAAGAAGATCTAGAACAAGAACAAAGTCAGGGTCAATTCTCTC 1539
QY 540 ACTCACATGGTGGTGGTAGTTGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAA 599
Db 1540 ACTCACATGGTGGTGGTAGTTGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAA 1599
QY 600 GAACAACCTTAAGGTATTGGGAGATCGATGGSCAAAACATCTGTAGATGGACAGAACCCGC 659
Db 1600 GAACAACCTTAAGGTATTGGGAGATCGATGGSCAAAACATCTGTAGATGGACAGAACCCGC 1659
QY 660 TGGGTTCTTTTACAAGACATCCTTCTCAAATGGCAACGCTCTTACTGAAGAACAGTGCCTT 719
Db 1660 TGGGTTCTTTTACAAGACAGTTCTTGACCACTGGAAGCGTCTGCACCTTTCTCTGCAGGAA 1719
QY 720 TTAGTGCATGGCTTTCAGAAAAAAGAGATGCAGTGAACAAGATTACACACAACCTGGCTTT 779
Db 1720 CTCTGGTGTGGCTACAGCTGAAGATGATGAATTAAGCCGGCAGGCACCTATTGGAGGC 1779
QY 780 AAAGATCAAAATGAAATGTTATCAAGTCTTCAAAAACCTGGCCGCTTTTAAAAGCGGATCTA 839
Db 1780 GACTTTCAGCAGTTTCAGAAGCAGAACGATGTACATAGGCGCTTCAAGAGGGAATTCAAA 1839
QY 840 GAAAAAGAAAAGCAATCCATGGGCAAACTGTA-----TTCACTCAAAACAAG 885
Db 1840 ACTAAGAACCTGTAAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAG 1899
QY 886 ATCTTCTTTCAACACTGAAGAATAAGTCAGTGACCCAGAAAGACGGAAGCATGGCTGGA-- 943
Db 1900 CCTTTGGAAGGACTAGAGAAACTCTACCAGAGAGCCAGAGAGCTGCCTCCTGAGGAGAGA 1959
QY 944 -----TAACTTTGCCCGGTGTTGGGATAATTTAGTCCAAA 978
Db 1960 GCCCAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAA 2019
QY 979 AACTTTGAA-----AAGAGTACAGCACAGACCCCTTGAAGACTC 1016
Db 2020 AAATTGAACCTGCACCTCGCTGACTGGCAGAGAAAATAGATGAGACCCCTTGAAGACTC 2079
QY 1017 CAGGAACCTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGATC 1076
Db 2080 CAGGAACCTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGATC 2139
QY 1077 AAGGGATCCTGGCAGCCCGTGGCGATCTCCTCAATTGACTCTCTCCAAGATCACCTCGAG 1136
Db 2140 AAGGGATCCTGGCAGCCCGTGGCGATCTCCTCAATTGACTCTCTCCAAGATCACCTCGAG 2199
QY 1137 AAAGTCAAGGCACCTTCGAGGAGAAAATTGCGCCTCTGAAAGAGAACGTCAGCCACGTCAAT 1196
Db 2200 AAAGTCAAGGCACCTTCGAGGAGAAAATTGCGCCTCTGAAAGAGAACGTCAGCCACGTCAAT 2259
QY 1197 GACCTTGCTCGCCAGCTTACCACCTTTGGGCATTCAGCTCTCACCCGTATAACCTCAGCACT 1256
Db 2260 GACCTTGCTCGCCAGCTTACCACCTTTGGGCATTCAGCTCTCACCCGTATAACCTCAGCACT 2319
QY 1257 CTGGAAGACCTGAACACACAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGG 1316
Db 2320 CTGGAAGACCTGAACACACAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGG 2379
QY 1317 CAGCTGCATGAAGCCCCACAGGACTTTGGTCCAGCATCTCAGCACTTTCTTCCACGTCT 1376
Db 2380 CAGCTGCATGAAGCCCCACAGGACTTTGGTCCAGCATCTCAGCACTTTCTTCCACGTCT 2439
QY 1377 GTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAAACAAAGTGCCCTACTATATCAACCAC 1436

Db 2440 GTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAAACAAAGTGCCCTACTATATCAACCAC 2499
QY 1437 GAGACTCAAAACAACACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCT 1496
Db 2500 GAGACTCAAAACAACACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCT 2559
QY 1497 GACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAG 1556
Db 2560 GACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAG 2619
QY 1557 AAGGCCCTTTGCTTGGATCTCTTGAAGCTGTCAGCTGCATGTGATGCCTTGGACCCAGC 1616
Db 2620 AAGGCCCTTTGCTTGGATCTCTTGAAGCTGTCAGCTGCATGTGATGCCTTGGACCCAGC 2679
QY 1617 AACCTCAAGCAAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTGACCACT 1676
Db 2680 AACCTCAAGCAAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTGACCACT 2739
QY 1677 ATTTATGACCGCCTGGAGCAAGAGCAGCAACAATTTGGTCAACGTCCTCTCTGCGTGGAT 1736
Db 2740 ATTTATGACCGCCTGGAGCAAGAGCAGCAACAATTTGGTCAACGTCCTCTCTGCGTGGAT 2799
QY 1737 ATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCGTGTC 1796
Db 2800 ATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCGTGTC 2859
QY 1797 CTGTCTTTTAAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAGA 1856
Db 2860 CTGTCTTTTAAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAGA 2919
QY 1857 TACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTGTGACCAGCGCAGGCTGGGCCTC 1916
Db 2920 TACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTGTGACCAGCGCAGGCTGGGCCTC 2979
QY 1917 CTTCTGCATGATTTCTATCCAAATTCGAAGACAGTTGGTGAAAGTTGCATCCTTTGGGGC 1976
Db 2980 CTTCTGCATGATTTCTATCCAAATTCGAAGACAGTTGGTGAAAGTTGCATCCTTTGGGGC 3039
QY 1977 AGTAACATTGAGCCAAAGTGTCCGGA 2001
Db 3040 AGTAACATTGAGCCAAAGTGTCCGGA 3064

RESULT 4
AAD37259
ID AAD37259 standard; DNA; 4476 BP.
XX
AC AAD37259;
XX
DT 21-AUG-2002 (first entry)
XX
DE Adeno-associated virus vector plasmid, AAV-MCK-3510.
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Unidentified.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
XX
PI Xiao X;
XX
DR WPI; 2002-049342/06.

XX New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -
XX
PS Example 1; Page 63-65; 71pp; English.
XX
CC The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is AAV
CC vector plasmid construct containing human dystrophin minigenes, a muscle
CC creatine kinase (MCK) promoter and a small polyA signal sequence.
XX
SQ Sequence 4476 BP; 1252 A; 1096 C; 1127 G; 1001 T; 0 other;

Query Match 79.6%; Score 1593; DB 24; Length 4476;
Best Local Similarity 87.7%; Pred. No. 0;
Matches 1811; Conservative 0; Mismatches 190; Indels 64; Gaps 4;

QY 1 GGCAGTTCATTGATGGAGAGTGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAA 60
Db 1756 GGCAGTTCATTGATGGAGAGTGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAA 1815
QY 61 GTATTATCGTGGCTTCTTTCTGCTGAGGACACATTGCAAGCACAAAGGAGAGATTCTTAAT 120
Db 1816 GTATTATCGTGGCTTCTTTCTGCTGAGGACACATTGCAAGCACAAAGGAGAGATTCTTAAT 1875
QY 121 GATCTGGAAGTGGTGAAGACCAGTTCATACACTCATGAGGGGTACATGATGGATTGACA 180
Db 1876 GATCTGGAAGTGGTGAAGACCAGTTCATACACTCATGAGGGGTACATGATGGATTGACA 1935
QY 181 GCCCATCAGGGCCGGGTTGGTAATATCTTACAAATGGGAAGTAAGCTGATTGGAACAGGA 240
Db 1936 GCCCATCAGGGCCGGGTTGGTAATATCTTACAAATGGGAAGTAAGCTGATTGGAACAGGA 1995
QY 241 AAATTATCAGAAGATGAAGAAACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAAGA 300
Db 1996 AAATTATCAGAAGATGAAGAAACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAAGA 2055
QY 301 TGGGAATGCCTCAGGGTAGCTAGCATGGAACAAACAAAGCAATTTACATAGAGTTTAAATG 360
Db 2056 TGGGAATGCCTCAGGGTAGCTAGCATGGAACAAACAAAGCAATTTACATAGAGTTTAAATG 2115
QY 361 GATCTCAGAATC-GAAACTGAAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGAA 419
Db 2116 GATCTCAGAATCAGAAACTGAAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGAA 2175
QY 420 ACAAGGAAATGGAGGAAGAGCCCTCTTGGACCTGATCTTGAAGACCTAAACGCCAAGTA 479
Db 2176 ACAAGGAAATGGAGGAAGAGCCCTCTTGGACCTGATCTTGAAGACCTAAACGCCAAGTA 2235
QY 480 CAACAACATAAGGTGCTTCAAGAAGATCTAGAACAAGAAACAAGTCAAGGTCAATTCTCTC 539
Db 2236 CAACAACATAAGGTGCTTCAAGAAGATCTAGAACAAGAAACAAGTCAAGGTCAATTCTCTC 2295
QY 540 ACTCACATGTTGGTGGTAGTTGATGAATCTACTGGAGATCACGCAACTGCTGCTTGGAA 599
Db 2296 ACTCACATGTTGGTGGTAGTTGATGAATCTACTGGAGATCACGCAACTGCTGCTTGGAA 2355
QY 600 GAACAACCTAAGGTATTGGGAGATCGATGGGCAACATCTGTAGATGGACAGAGACCCG 659
Db 2356 GAACAACCTAAGGTATTGGGAGATCGATGGGCAACATCTGTAGATGGACAGAGACCCG 2415
QY 660 TGGGTTCTTTTACAAGACATCCTTCTCAATGGCAACGTCCTTACTGAAGAAGACAGTGCCTT 719
Db 3496 ATTTATGACCGCCTGGAGCAAGAGCACAAACAATTTGGTCAACGTCCTCTCTCGCTGGAT 3555

Db 2416 TGGGTTCTTTTACAAGACAGTTCTGACCAGTGGAAAGCGTCTGCACCTTTCTCTGCAGGAA 2475
QY 720 TTTAGTGCATGGCTTTTCAGAAAAAAGAGATGCAGTGAACAAGATTTCACACAACCTGGCTTT 779
Db 2476 CTTCTGGTGTGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACTATTGGAGGC 2535
QY 780 AAAGATCAAAATGAAATGTTATCAAGTCTTCAAAAACTGGCCGTTTTAAAAAGCGGATCTA 839
Db 2536 GACTTTCAGCAGTTCAGAAGCAGCAACGATGTACATAGGCGCTTCAAGAGGGAATTGAAA 2595
QY 840 GAAAAAGAAAAAGCAATCCATCCATGGGCAAACTGTA-----TTCACTCAAAACAAG 885
Db 2596 ACTAAAGAACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAG 2655
QY 886 ATCTTCTTTCAACACTGAAGAATAAGTCACTGACCCAGAACGCGGAGCATGGCTGGA-- 943
Db 2656 CTTTGGAAAGGACTAGAGAAACTCTACCAGGAGCCAGAGAGCTGCCCTCTGAGGAGAGA 2715
QY 944 -----TAACTTTGCCCGGTGTTGGGATAAATTTAGTCCAAA 978
Db 2716 GCCCAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGAA 2775
QY 979 AACTTGAA-----AAGAGTACAGCACAGACCCCTTGAAGACTC 1016
Db 2776 AAATTGAACCTGCACCTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAGACTC 2835
QY 1017 CAGGAACCTTCAAGAGGCCACGGATGAGTGGACCTCAAGCTGCGCCAAAGTGAGGTGATC 1076
Db 2836 CAGGAACCTTCAAGAGGCCACGGATGAGTGGACCTCAAGCTGCGCCAAAGTGAGGTGATC 2895
QY 1077 AAGGATCCTGGCAGCCCGTGGCGGATCTCCTCATCTCTCCTCAAGATCACCTCGAG 1136
Db 2896 AAGGATCCTGGCAGCCCGTGGCGGATCTCCTCATCTCTCCTCAAGATCACCTCGAG 2955
QY 1137 AAAGTCAAGGCACCTTCGAGGAGAAATTCGCGCTCTGAAAGAGAACGCTGAGCCACGTCAAT 1196
Db 2956 AAAGTCAAGGCACCTTCGAGGAGAAATTCGCGCTCTGAAAGAGAACGCTGAGCCACGTCAAT 3015
QY 1197 GACCTTGCTGCCAGCTTACCACCTTTGGGCACTTCCACCGTATACCTCAGCAGCT 1256
Db 3016 GACCTTGCTGCCAGCTTACCACCTTTGGGCACTTCCACCGTATACCTCAGCAGCT 3075
QY 1257 CTGGAAGACCTGAACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGAGCCGAGTCAGG 1316
Db 3076 CTGGAAGACCTGAACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGAGCCGAGTCAGG 3135
QY 1317 CAGTGCATGAAGCCCAACAGGAGCTTTGGTCCAGCATCTCAGCAGCTTTTCTTCCAGCTCT 1376
Db 3136 CAGTGCATGAAGCCCAACAGGAGCTTTGGTCCAGCATCTCAGCAGCTTTTCTTCCAGCTCT 3195
QY 1377 GTCCAGGTCCCTGGGAGAGAGCCATCTCGCCAAAACAAAAGTGCCTACTATATCAACCAC 1436
Db 3196 GTCCAGGTCCCTGGGAGAGAGCCATCTCGCCAAAACAAAAGTGCCTACTATATCAACCAC 3255
QY 1437 GAGACTCAAAACAACTTGTGGGACCATCCCAAAAATGACAGAGCTCTACCAGTCTTAGCT 1496
Db 3256 GAGACTCAAAACAACTTGTGGGACCATCCCAAAAATGACAGAGCTCTACCAGTCTTAGCT 3315
QY 1497 GACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAAGACTGCAG 1556
Db 3316 GACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAAGACTGCAG 3375
QY 1557 AAGGCCCTTTGCTTGGATCTCTTGAGCCCTGTGAGTGCATGTGATGCCCTTGGACCAGCAC 1616
Db 3376 AAGGCCCTTTGCTTGGATCTCTTGAGCCCTGTGAGTGCATGTGATGCCCTTGGACCAGCAC 3435
QY 1617 AACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTGACCACT 1676
Db 3436 AACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTGACCACT 3495
QY 1677 ATTTATGACCGCCTGGAGCAAGAGCACAAACAATTTGGTCAACGTCCTCTCTCGCTGGAT 1736
Db 3496 ATTTATGACCGCCTGGAGCAAGAGCACAAACAATTTGGTCAACGTCCTCTCTCGCTGGAT 3555

1737	ATGTGCTCTGAAC	TGGCTGCTGAATGTTATGATACGGACGAACAGGGAGGATCCGTGTC	1796
1738			
3556	ATGTGCTCTGAAC	TGGCTGCTGAATGTTATGATACGGACGAACAGGGAGGATCCGTGTC	3615
1797	CTGTCTTTTAAAC	TGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAGA	1856
3616	CTGTCTTTTAAAC	TGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAGA	3675
1857	TACCTTTTCAAGCA	AGTGGCAAGTTCAACACAGGATTTTGACACCAGCGCAGGCTGGGCCTC	1916
3676	TACCTTTTCAAGCA	AGTGGCAAGTTTCAACACAGGATTTTGACACCAGCGCAGGCTGGGCCTC	3735
1917	CTTCTGCATGAT	TCTATCCAAATTTCCAAGACAGTTGGGTGAAGTTGCATCCTTTGGGGGC	1976
3736	CTTCTGCATGAT	TCTATCCAAATTTCCAAGACAGTTGGGTGAAGTTGCATCCTTTGGGGGC	3795
1977	AGTAACATTGAGC	CAAGTGTCCGGA	2001
3796	AGTAACATTGAGC	CAAGTGTCCGGA	3820

F.S.II.T. 5

A748568

2 AAZ48568 standard: cDNA to mRNA: 4402 BP.

AAZ48568:

31-MAR-2000 (first entry)

A rod shortened dystrophin (deltaDysAH3) encoding nucleotide sequence.

Muscular dystrophy; rod domain; adeno-associated virus; AAV;

dystrophin gene; truncated; ds:

S Homo sapiens.

JP11318467-A.

24 - NOV - 1999.

08-MAY-1998; 98JP-0142134.

08-MAY-1998: 98JP-0142134.

(KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
(KOKU-) KOKURITSU SEISHIN SHINKEI CENT SOCHO.

WPI; 2000-100771/09.

P-PSDB; AAY59239.

.. T T
A gene encoding a shortened dystrophin - useful for the treatment of muscular dystrophy T

Claim 7: Page 21-22: 4pp: Japanese.

The invention provides a gene for the treatment of muscular dystrophy having at least one rod repeat structure of hinge 1, hinge 4 and rod domain of dystrophin gene and having a base sequence of 4.5 kb. The gene and a gene-introducing medium consisting of an adeno-associated virus (AAV) vector or lentivirus vector containing the rod shortened dystrophin genes can be used for the genetic treatment of muscular dystrophy of low immune reaction. The present sequence represents a rod shortened dystrophin encoding sequence.

Sequence 4402 BP: 1339 A: 984 C: 1010 G: 1069 T: 0 other:

Query Match	66.6%	Score 1332;	DB 21;	Length 4402;
Best Local Similarity	83.8%	Pred. NO. 0;		
Matches 1677: Conservative	0;	Mismatches	0;	Indels 325;
				Gaps 2;

1 GGCAGTTCATTGATGGAGAGTGAAGTAAACCTGGACCGTTATCAAACAGCTTTAGAAGAA 60

Db	1195	GGCAGTTCATTGATGGAGAGTGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAGAA	1254
QY	61	GTATTATCGTGGCTTCTTTCTGCTGAGGACACATTGCAAGCACAAAGGAGAGATTTCTAAT	120
Db	1255	GTATTATCGTGGCTTCTTTCTGCTGAGGACACATTGCAAGCACAAAGGAGAGATTTCTAAT	1314
QY	121	GATGTGGAAGTGGTGAAGAACACCAGTTTCATACTCATGAGGGGTACATGATGGATTTTGACA	180
Db	1315	GATGTGGAAGTGGTGAAGAACACCAGTTTCATACTCATGAGGGGTACATGATGGATTTTGACA	1374
QY	181	GCCCATCAGGGCCCGGTTGGTAATAATTCTACAATTGGGAAGTAAAGCTGATTGGAACAGGA	240
Db	1375	GCCCATCAGGGCCCGGTTGGTAATAATTCTACAATTGGGAAGTAAAGCTGATTGGAACAGGA	1434
QY	241	AAATTATCAGAAGATGAAGAAACTGAAGTACAAGAGCAGATGAATCTCCTAAATTTCAAGA	300
Db	1435	AAATTATCAGAAGATGAAGAAACTGAAGTACAAGAGCAGATGAATCTCCTAAATTTCAAGA	1494
QY	301	TGGGAATGCCTCAGGGTAGCTAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATG	360
Db	1495	TGGGAATGCCTCAGGGTAGCTAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATG	1554
QY	361	GATCTCCAGAATC-GAAACTGAAAAGAGTTGAATGACTGGCTAAACAAAAACAGAAAGA	419
Db	1555	GATCTCCAGAATCAGAAAACTGAAAGAGTTGAATGACTGGCTAAACAAAAACAGAAAGA	1614
QY	420	ACAAGGAAAAATGGAGGAAGAGCCTCTTTGGACCTGATCTTGAAGACCTAAAACGCCAAGTA	479
Db	1615	ACAAGGAAAAATGGAGGAAGAGCCTCTTTGGACCTGATCTTGAAGACCTAAAACGCCAAGTA	1674
QY	480	CAACAACATAAGGTGCTTCAAGAAAGATCTAGAACAAAGAACAAAGTCAGGGTCAATTCTCTC	539
Db	1675	CAACAACATAAGGTGCTTCAAGAAAGATCTAGAACAAAGAACAAAGTCAGGGTCAATTCTCTC	1734
QY	540	ACTCACATGGTGGTAGTTGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAA	599
Db	1735	ACTCACATGGTGGTAGTTGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAA	1794
QY	600	GAACAACTTAAGGTATTTGGGAGATCGATGGGCAAAACATCTGTAGATGGACAGAAAGACCCG	659
Db	1795	GAACAACTTAAGGTATTTGGGAGATCGATGGGCAAAACATCTGTAGATGGACAGAAAGACCCG	1854
QY	660	TGGGTTCTTTTACAAGACATCCTTCTCAAATGGCAACGTCTTACTGAAGAACAGTGCCTT	719
Db	1855	TGGGTTCTTTTACAAGACATCCTTCTCAAATGGCAACGTCTTACTGAAGAACAGTGCCTT	1914
QY	720	TTTAGTGCATGGCTTTCAGAAAAAAGAAGATGCAGTGAACAAGATTACACAACTGGCTTT	779
Db	1915	TTTAGTGCATGGCTTTCAGAAAAAAGAAGATGCAGTGAACAAGATTACACAACTGGCTTT	1974
QY	780	AAAGATCAAAATGAAATGTTATCAAGTCTTCAAAAACCTGGCCGTTTAAAAGCGGATCTA	839
Db	1975	AAAGATCAAAATGAAATGTTATCAAGT-----	2001
QY	840	GAAGAAGAAAAAGCAATCCATGGGCAAACTGTATTCACTCAAAACAAGATCTTCTTTCAACA	899
Db	2002	-----	2001
QY	900	CTGAAGAAATAAGTCAGTGACCCAGAAAGACGGAAGCATGGCTGGATAAATTTGCCCGGTGT	959
Db	2002	-----	2001
QY	960	TGGGATAATTAGTCCAAAAAATTGAAAAAGATTACAGCACAGACCCCTTGAAGAAGACTCCAG	1019
Db	2002	-----	2001
QY	1020	GAACTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCAAGCTGAGGTGATCAAG	1079
Db	2002	-----	2001
QY	1080	GGATCCTGGCAGCCCGTGGGCGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGAGAAA	1139
Db	2002	-----	2001


```
QY 1140 GTCAAGGCACTTCGAGGAGAAATTCGCCCTCTGAAAGAGAACGTGAGCCACGTCATGAC 1199
    |||
Db 2011 GTCAAGGCACTTCGAGGAGAAATTCGCCCTCTGAAAGAGAACGTGAGCCACGTCATGAC 2070
    |||
QY 1200 CTTGCTCGCCAGCTTACACACTTTGGGCATTCAGCTCTCACCCGTATAACCTCAGCACTCTG 1259
    |||
Db 2071 CTTGCTCGCCAGCTTACACACTTTGGGCATTCAGCTCTCACCCGTATAACCTCAGCACTCTG 2130
    |||
QY 1260 GAAGACCTGAACACACAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGGCAG 1319
    |||
Db 2131 GAAGACCTGAACACACAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGGCAG 2190
    |||
QY 1320 CTGCATGAAGCCACAGGCACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGTCCTGTC 1379
    |||
Db 2191 CTGCATGAAGCCACAGGCACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGTCCTGTC 2250
    |||
QY 1380 CAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCACGAG 1439
    |||
Db 2251 CAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTATATCAACCACGAG 2310
    |||
QY 1440 ACTCAAAACAACTTCTGCGGACCATCTCGCCAAACAAAGTGCCCTACTATATCAACCACGAG 1499
    |||
Db 2311 ACTCAAAACAACTTCTGCGGACCATCTCGCCAAACAAAGTGCCCTACTATATCAACCACGAG 2370
    |||
QY 1500 CTGAATAATGTCAGATCTCAGCTTATAGGACTGCCATGAAACTCCGGAAGACTGCAGAAG 1559
    |||
Db 2371 CTGAATAATGTCAGATCTCAGCTTATAGGACTGCCATGAAACTCCGGAAGACTGCAGAAG 2430
    |||
QY 1560 GCCCTTTGCTGGATCTCTTGAGCCTGTGAGCTGCATGTCATGCCCTTGGACCAGCACAAAC 1619
    |||
Db 2431 GCCCTTTGCTGGATCTCTTGAGCCTGTGAGCTGCATGTCATGCCCTTGGACCAGCACAAAC 2490
    |||
QY 1620 CTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTTAATTTGTTGACCACTATT 1679
    |||
Db 2491 CTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTTAATTTGTTGACCACTATT 2550
    |||
QY 1680 TATGACCGCCTGGAGCAAGAGCACAAATTTGGTCAACGTCCCTCTCTGCGTGGATATG 1739
    |||
Db 2551 TATGACCGCCTGGAGCAAGAGCACAAATTTGGTCAACGTCCCTCTCTGCGTGGATATG 2610
    |||
QY 1740 TGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACCAACAGGGAGGATCCGTGTCCTG 1799
    |||
Db 2611 TGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGACCAACAGGGAGGATCCGTGTCCTG 2670
    |||
QY 1800 TCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAAGTACAGATAC 1859
    |||
Db 2671 TCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAAGTACAGATAC 2730
    |||
QY 1860 CTTTTCAGCAAGTGGCAAGTTTCAACAGGATTTGTGACCAGCGCAGGCTGGCCCTCCTT 1919
    |||
Db 2731 CTTTTCAGCAAGTGGCAAGTTTCAACAGGATTTGTGACCAGCGCAGGCTGGCCCTCCTT 2790
    |||
QY 1920 CTGCATGATTTATCCAAATTTCCAAGACAGTTGGGTGAAGTTGCATCCTTTGGGGGCGAGT 1979
    |||
Db 2791 CTGCATGATTTATCCAAATTTCCAAGACAGTTGGGTGAAGTTGCATCCTTTGGGGGCGAGT 2850
    |||
QY 1980 AACATTGAGCCCAAGTGTCGGGA 2001
    |||
Db 2851 AACATTGAGCCCAAGTGTCGGGA 2872
    |||
```

```
RESULT 6
AAZ48567
ID ID
AAZ48567 standard; cDNA to mRNA; 4402 BP.
XX AC
AAZ48567;
AC AC
XX AC
31-MAR-2000 (first entry)
XX AC
XX AC
A rod shortened dystrophin (deltaDysAx11) encoding nucleotide sequence.
DE XX
Muscular dystrophy; rod domain; adeno-associated virus; AAV;
KW XX
```

```
KW dystrophin gene; truncated; ds.
XX Homo sapiens.
OS JP11318467-A.
PN 24-NOV-1999.
XX 08-MAY-1998; 98JP-0142134.
PD 08-MAY-1998; 98JP-0142134.
XX (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.
PA (KOKU-) KOKURITSU SEISHIN SHINKEI CENT SOCHO.
XX WPI; 2000-100771/09.
DR P-PSDB; AAY59238.
XX
```

A gene encoding a shortened dystrophin - useful for the treatment of muscular dystrophy

Claim 7; Page 16-17; 44pp; Japanese.

The invention provides a gene for the treatment of muscular dystrophy having at least one rod repeat structure of hinge 1, hinge 4 and rod domain of dystrophin gene and having a base sequence of 4.5 kb. The gene and a gene-introducing medium consisting of an adeno-associated virus (AAV) vector or lentivirus vector containing the rod shortened dystrophin genes can be used for the genetic treatment of muscular dystrophy of low immune reaction. The present sequence represents a rod shortened dystrophin encoding sequence.

Sequence 4402 BP; 1329 A; 1000 C; 1019 G; 1054 T; 0 other;

Query Match 63.4%; Score 1269.2; DB 21; Length 4402;

Best Local Similarity 82.1%; Pred. No. 0;

Matches 1644; Conservative 0; Mismatches 33; Indels 325; Gaps 3;

```
QY 1 GGCAGTTCATTGATGGAGAGTGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAGAA 60
    |||
Db 1195 GGCAGTTCATTGATGGAGAGTGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAGAA 1254
    |||
QY 61 GTATTATCGTGGCTTCTTTCTGCTGAGGACACATTTGCAAGCACAAAGGAGAGATTTCTAAT 120
    |||
Db 1255 GTATTATCGTGGCTTCTTTCTGCTGAGGACACATTTGCAAGCACAAAGGAGAGATTTCTAAT 1314
    |||
QY 121 GATGTGGAAGTGGTGAAGACCAAGTTCATFACCTCATGAGGGGTACATGATGGATTTGACA 180
    |||
Db 1315 GATGTGGAAGTGGTGAAGACCAAGTTCATFACCTCATGAGGGGTACATGATGGATTTGACA 1374
    |||
QY 181 GCCATCAGGCGCGGTTGGTAATATTTACAATTTGGGAAGTAAGCTGATTTGGAACAGGA 240
    |||
Db 1375 GCCATCAGGCGCGGTTGGTAATATTTACAATTTGGGAAGTAAGCTGATTTGGAACAGGA 1434
    |||
QY 241 AAATTATCAGAAGATGAAGAACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGA 300
    |||
Db 1435 AAATTATCAGAAGATGAAGAACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGA 1494
    |||
QY 301 TGGCAATGCCTCAGGGTAGCTAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAATG 360
    |||
Db 1495 TGGCAATGCCTCAGGGTAGCTAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAATG 1554
    |||
QY 361 GATCTCCAGAAATC-GAAACTGAAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGAA 419
    |||
Db 1555 GATCTCCAGAAATCAGAAACTGAAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGAA 1614
    |||
QY 420 ACAAGGAAAAATGGAGGAAGAGCCCTTTGGACCTGATCTTTGAAGACCTAAACGCCAAGTA 479
    |||
Db 1615 ACAAGGAAAAATGGAGGAAGAGCCCTTTGGACCTGATCTTTGAAGACCTAAACGCCAAGTA 1674
    |||
QY 480 CAACAACATAAGGTGCTTCAAGAGAGATCTAGAACAAAGCAAGTCAGGGTCAATTCTCTC 539
    |||
Db 1675 CAACAACATAAGGTGCTTCAAGAGAGATCTAGAACAAAGCAAGTCAGGGTCAATTCTCTC 1734
    |||
```


QY 540 ACTCACATGGTGGTAGTTGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAA 599
Dbb 1735 ACTCACATGGTGGTAGTTGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAA 1794
QY 600 GAACAACTTAAGGTATTGGGAGATCGATGGGCAAAACATCTGTAGATGGACAGAAGACCGC 659
Dbb 1795 GAACAACTTAAGGAGGTCAATACTGAGTGGGAAAA----- 1829
QY 660 TGGGTTCTTTTACAAGACATCCTTCTCAAATGGCAACGCTTACTGAAGAACAGTGCCTT 719
Dbb 1830 ----- 1829
QY 720 TTTAGTCATGGCTTTTCAGAAAAAAGAGATGSCAGTGAACAAGATTACACAACTGGCCTT 779
Dbb 1830 ----- 1829
QY 780 AAAGATCAAAATGAAATGTTATCAAGTCTTCAAAAACTGGCCGTTTTTAAAGCGGATCTA 839
Dbb 1830 ----- 1829
QY 840 GAAAAAAGAAAGCAATCCATGGGCAAACTGTATTCACTCAACAAGATCTTCTTTCAACA 899
Dbb 1830 ----- 1829
QY 900 CTGAAGAATAAGTCAGTGACCCAGACGGAAGCATGGCTGGATAAATTTGCCCGGTGT 959
Dbb 1830 -----ATTGAACCTGCACCTCCGCTGACTGGCAGAGAAA----- 1862
QY 960 TGGGATAATTTAGTCCAAAAAATTTGAAAAGAGTACAGCACAGACCCCTTGAAAAGACTCCAG 1019
Dbb 1863 -----AATAGATGAGACCCCTTGAAAAGACTCCAG 1890
QY 1020 GAACTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTCGGCCAAGCTGAGGTGATCAAG 1079
Dbb 1891 GAACTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTCGGCCAAGCTGAGGTGATCAAG 1950
QY 1080 GGATCCTGGCAGCCCGTGGGCGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGAGAAA 1139
Dbb 1951 GGATCCTGGCAGCCCGTGGGCGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGAGAAA 2010
QY 1140 GTCAAGGCACTTCGAGGAGAAATTTGGCCTCTGAAAGAGAACGCTGAGCCACGTCATGAC 1199
Dbb 2011 GTCAAGGCACTTCGAGGAGAAATTTGGCCTCTGAAAGAGAACGCTGAGCCACGTCATGAC 2070
QY 1200 CTTGCTGCCAGCTTACCACCTTTGGSCATTCAGCTCTCACCGTATAAACCCTCAGCACTCTG 1259
Dbb 2071 CTTGCTGCCAGCTTACCACCTTTGGSCATTCAGCTCTCACCGTATAAACCCTCAGCACTCTG 2130
QY 1260 GAAGACCTGAACACCCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGGCAG 1319
Dbb 2131 GAAGACCTGAACACCCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGGCAG 2190
QY 1320 CTGCATGAAGCCCAACAGGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGTCGTGTC 1379
Dbb 2191 CTGCATGAAGCCCAACAGGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGTCGTGTC 2250
QY 1380 CAGGGTCCCTGGGAGAGAGCCATCTCGCCAAAAAAGTGCCCTACTATATCAACACAGAG 1439
Dbb 2251 CAGGGTCCCTGGGAGAGAGCCATCTCGCCAAAAAAGTGCCCTACTATATCAACACAGAG 2310
QY 1440 ACTCAAAACAACCTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGAC 1499
Dbb 2311 ACTCAAAACAACCTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGAC 2370
QY 1500 CTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAAG 1559
Dbb 2371 CTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAAG 2430
QY 1560 GCCCTTTGCTTGGATCTCTTGGCCTGTGAGCTGCATGTGATGCCTTGGACAGCACAAAC 1619
Dbb 2431 GCCCTTTGCTTGGATCTCTTGGCCTGTGAGCTGCATGTGATGCCTTGGACAGCACAAAC 2490

QY 1620 CTCAAGCAAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTTGTTGACCACATATT 1679
Dbb 2491 CTCAAGCAAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTTGTTGACCACATATT 2550
QY 1680 TATGACCGCCTGGAGCAAGAGCACACAATTTGGTCAACGTCCTCTCTGCGTGGATATG 1739
Dbb 2551 TATGACCGCCTGGAGCAAGAGCACACAATTTGGTCAACGTCCTCTCTGCGTGGATATG 2610
QY 1740 TGTCTGAACTGGCTGCTGAATGTTTATGATACGGGACGAACAGGAGGATCCGTGTCCTG 1799
Dbb 2611 TGTCTGAACTGGCTGCTGAATGTTTATGATACGGGACGAACAGGAGGATCCGTGTCCTG 2670
QY 1800 TCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAGATAC 1859
Dbb 2671 TCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTACAGATAC 2730
QY 1860 CTTTTCAGCAAGTGGCAAGTTCAACAGGATTTGTGACCGCAGCGCTGGGCCCTCCTT 1919
Dbb 2731 CTTTTCAGCAAGTGGCAAGTTCAACAGGATTTGTGACCGCAGCGCTGGGCCCTCCTT 2790
QY 1920 CTGCATGATCTATCCAAATTCCAAGACAGTTGGGTGAAAGTTGCATCCTTTGGGGGCGAGT 1979
Dbb 2791 CTGCATGATCTATCCAAATTCCAAGACAGTTGGGTGAAAGTTGCATCCTTTGGGGGCGAGT 2850
QY 1980 AACATTGAGCCCAAGTGTCCGGA 2001
Dbb 2851 AACATTGAGCCCAAGTGTCCGGA 2872
RESULT 7
AAD37237
ID AAD37237 standard; DNA; 3858 BP.
XX
AC AAD37237;
XX
DT 21-AUG-2002 (first entry)
XX
DE Human dystrophin minigene delta3849.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Homo sapiens.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
XX
PI Xiao X;
XX
DR WPI; 2002-049342/06.
XX
PT New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -
XX
PS Example 1; Page 48-49; 71pp; English.
XX
CC The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression

control element, in a recombinant adeno-associated virus or retrovirus is useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular dystrophy (BMD) in a mammalian subject. The present sequence is human dystrophin minigene delta3849 containing nucleotides 1-1668 (N-terminus, hinge H1 and rods R1, R2), 8059-10227 (rods R22, R23 and R24, hinge H4 and CR domain) and 11047-11058 (dystrophin last 3 amino acids).

Db	2597	CTCCAAGGTGAAATTGAAGCTCACACAGATGTTTATCACAACTGGATGAAACAGCCAA	2656
QY	885	GATCTTCTTTCAACACTGAAGAATAAGTCAGTGACCCAGAAAGACGGAAGCATGGCTGGAT	944
Db	2657	AAATCCTGAGATCCCTGGAAGGTTCCGATGATGCAGTCTGTTACAAAGACGTTTGAT	2716
QY	945	AACCTTGCCCGGTGTGGATAAATTAGTCCAAAAACTTGAAAAAGAGTACAGCAC	999
Db	2717	AACATGAACCTTCAAGTGGAGTGAACCTCGGAAAAAGTCTCTCAACATTAGGTCCCAATTG	2776
QY	1000	-----	999
Db	2777	GAAGCCAGTTCTGACCAGTGGAAAGCGTCTGCACCTTTCTCTGCAGGAACTTCTGTTGTGG	2836
QY	1000	-----	999
Db	2837	CTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCTATTGGAGGGGACTTTTCCAGCA	2896
QY	1000	-----	999
Db	2897	GTTCAGAAGCAGAACGATGTACATAGGGCCTTCAAGAGGGAATTGAAAACTAAAGAACCT	2956
QY	1000	-----	999
Db	2957	GTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAAGGA	3016
QY	1000	-----	999
Db	3017	CTAGAGAAACTCTACCAGGAGCCCAAGAGAGCTGCCTCCTGAGGAGAGAGCCAGAAATGTC	3076
QY	1000	-----	999
Db	3077	ACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTAGTGGGAAAAAATTGAACCTG	3136
QY	1000	-----AGACCTTTGAAAGACTCCAGGAACCTTCAA	1028
Db	3137	CACCTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTTGAAAGACTCCAGGAACCTTCAA	3196
QY	1029	GAGGCCACGGATGAGCTGGACCTCAAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATCCTGG	1088
Db	3197	GAGGCCACGGATGAGCTGGACCTCAAAGCTGCGCCAAAGCTGAGGTGATCAAGGGATCCTGG	3256
QY	1089	CAGCCCGTGGCGATCTCCTCATTGACTCTCTCCAAGATCACCTCGAGAAAAGTCAAGGCA	1148
Db	3257	CAGCCCGTGGCGATCTCCTCATTGACTCTCCAAGATCACCTCGAGAAAAGTCAAGGCA	3316
QY	1149	CTTCGAGGAGAAAATTGCGCCTCTGAAAGAGAAAGCTGAGCCACGTCAATGACCTTGCTCGC	1208
Db	3317	CTTCGAGGAGAAAATTGCGCCTCTGAAAGAGAAAGCTGAGCCACGTCAATGACCTTGCTCGC	3376
QY	1209	CAGCTTACCACCTTTGGGCATTCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCTG	1268
Db	3377	CAGCTTACCACCTTTGGGCATTCAGCTCTCACCGTATAACCTCAGCACTCTGGAAGACCTG	3436
QY	1269	AACACCAGATGGAAGCTTCTGCAGGTGGCGTCTGAGGACCGAGTCAGGCAGCTGCATGAA	1328
Db	3437	AACACCAGATGGAAGCTTCTGCAGGTGGCGTCTGAGGACCGAGTCAGGCAGCTGCATGAA	3496
QY	1329	GCCACACAGGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGTCTGTCCAGGGTCCC	1388
Db	3497	GCCACACAGGACTTTGGTCCAGCATCTCAGCACTTTCTTTCCACGTCTGTCCAGGGTCCC	3556
QY	1389	TGGGAGAGAGCCATCTCGCCAAAACAAAGTGCCCTACTATATCAACCCACGAGACTCAAAACA	1448
Db	3557	TGGGAGAGAGCCATCTCGCCAAAACAAAGTGCCCTACTATATCAACCCACGAGACTCAAAACA	3616
QY	1449	ACTTGCTGGGACCATCCCCAAAATGACAGAGCTCTACCACTTTTAGCTGACCTGAATAAT	1508
Db	3617	ACTTGCTGGGACCATCCCCAAAATGACAGAGCTCTACCACTTTTAGCTGACCTGAATAAT	3676
QY	1509	GTCAGATTCTCAGCTTATAGGACTGCCATGAACTCCGAAGACTGCAGAAAGGCCCTTTGC	1568

Db	3677	GT	CAGAT	TCT	CAGCTT	TATAGG	ACTGCC	ATGAA	ACTCCG	AAGACT	GCAG	AAGGCC	CTTTGC	3736
QY	1569	TT	GGAT	CTCT	TGAG	CCTGT	CAGCT	GCAT	GTGAT	GCCTTG	GACCA	GCACA	ACCTCAAGCAA	1628
Db	3737	TT	GGAT	CTCT	TGAG	CCTGT	CAGCT	GCAT	GTGAT	GCCTTG	GACCA	GCACA	ACCTCAAGCAA	3796
QY	1629	AA	TACC	AGCC	CCAT	GGATA	TCTC	CTGC	AGATT	ATTAAT	TGTT	TGAC	CACTATTTATGACCGC	1688
Db	3797	AA	TACC	AGCC	CCAT	GGATA	TCTC	CTGC	AGATT	ATTAAT	TGTT	TGAC	CACTATTTATGACCGC	3856
QY	1689	CT	GGAG	CAAG	AGCACA	CAAA	TTTGG	TCAAC	GTCCCT	CTCT	TGCG	TGGAT	ATGTGTGAAC	1748
Db	3857	CT	GGAG	CAAG	AGCACA	CAAA	TTTGG	TCAAC	GTCCCT	CTCT	TGCG	TGGAT	ATGTGTGAAC	3916
QY	1749	TG	GCTG	CTGA	ATGTTT	TATG	ATAC	GGGAC	GAAC	AGGAG	GATCC	GTGTCTCTTTTAAA	1808	
Db	3917	TG	GCTG	CTGA	ATGTTT	TATG	ATAC	GGGAC	GAAC	AGGAG	GATCC	GTGTCTCTTTTAAA	3976	
QY	1809	ACT	GGCAT	CA	TTTCC	CTGT	GTAA	AGCAC	ATTTG	GAGACA	AGTAC	AGATA	ACCTTTTCAAG	1868
Db	3977	ACT	GGCAT	CA	TTTCC	CTGT	GTAA	AGCAC	ATTTG	GAGACA	AGTAC	AGATA	ACCTTTTCAAG	4036
QY	1869	CA	AGTGG	CAAG	TTCA	ACAGG	ATTTT	TGTG	ACCAG	CGCG	AGGCT	TGGG	CTCTCTGCA	1928
Db	4037	CA	AGTGG	CAAG	TTCA	ACAGG	ATTTT	TGTG	ACCAG	CGCG	AGGCT	TGGG	CTCTCTGCA	4096
QY	1929	TC	TATCC	AAAT	TTCC	AAGAC	AGTTGG	GGTGA	AGTTG	CA	TCTT	TGGGG	GCAGTAACATTGAG	1988
Db	4097	TC	TATCC	AAAT	TTCC	AAGAC	AGTTGG	GGTGA	AGTTG	CA	TCTT	TGGGG	GCAGTAACATTGAG	4156
QY	1989	CCA	AGTGT	CCGGA	2001									
Db	4157	CCA	AGTGT	CCGGA	4169									

RESULT 9
AAD37263

AAD37203
ID AAD37263 standard; DNA: 4848 BP.

AAD37263;
 21-AUG-2002 (first entry)
 Adeno-associated virus (AAV) vector plasmid, AAV-CMV-delta3849.
 Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
 adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
 Becker muscular dystrophy; ds.
 Chimeric - Homo sapiens.
 Chimeric - Cytomegalovirus.
 Chimeric - Unidentified.
 WO200183695-A2.
 08-NOV-2001.
 27-APR-2001; 2001WO-US13677.
 28-APR-2000; 2000US-200777P.
 (XIAO/) XIAO X.
 Xiao X;
 WPI; 2002-049342/06.
 New dystrophin minigene for treating Duchenne or Becker muscular
 dystrophy comprises an N-terminal domain or modified N-terminal domain
 rod repeats, H1 and H4 domains and a cysteine rich domain of a
 dystrophin gene -
 Example 1; Page 68-70; 71pp; English.

XX The present invention relates to an isolated nucleotide sequence encoding
:C a dystrophin minigene. The minigene comprises N-terminal or modified
:C N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
:C domains and cysteine-rich domains of dystrophin or utrophin genes. The
:C invention also relates to a recombinant adeno-associated virus (AAV)
:C comprising dystrophin minigene operably linked to an expression control
:C element. The dystrophin minigene in operable linkage with an expression
:C control element, in a recombinant adeno-associated virus or retrovirus is
:C useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
:C dystrophy (BMD) in a mammalian subject. The present sequence is AAV
:C vector plasmid construct containing human dystrophin minigenes, a
:C cytomegalovirus (CMV) promoter and a small polyA signal sequence.
XX
:Q Sequence 4848 BP; 1405 A; 1144 C; 1180 G; 1119 T; 0 other;

Query Match 63.1%; Score 1263; DB 24; Length 4848;
Best Local Similarity 75.3%; Pred. No. 0;
Matches 1816; Conservative 0; Mismatches 185; Indels 412; Gaps 3;

1 GGCAGTTTCATGAGAGAGTGAAGTAAACCTGGACCGTTATCAACACAGCTTTAGAAAGAA 60
1780 GGCAGTTTCATGAGAGAGTGAAGTAAACCTGGACCGTTATCAACACAGCTTTAGAAAGAA 1839
61 GTATTATCGTGGCTTCTTTCTGCTGAGGACACATTTGCAAGCACAAAGGAGAGATTTCTAAT 120
1840 GTATTATCGTGGCTTCTTTCTGCTGAGGACACATTTGCAAGCACAAAGGAGAGATTTCTAAT 1899
121 GATGTGGAAGTGGTGAAGACACAGTTTCATCTACTCATGAGGGGTACATGATGGATTGACA 180
1900 GATGTGGAAGTGGTGAAGACACAGTTTCATCTACTCATGAGGGGTACATGATGGATTGACA 1959
181 GCCCATCAGGGCCGGGTTGGTGAATATTTCTACAATTTGGGAAGTAAGCTGATTGGAACACGGA 240
1960 GCCCATCAGGGCCGGGTTGGTGAATATTTCTACAATTTGGGAAGTAAGCTGATTGGAACACGGA 2019
241 AAATTATCAGAAGATGAAGAAACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCGAAGA 300
2020 AAATTATCAGAAGATGAAGAAACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCGAAGA 2079
301 TGGGAATGCCTCAGGGTAGCTAGCATGGAAGAAACAAAGCAATTTACATAGAGTTTAAATG 360
2080 TGGGAATGCCTCAGGGTAGCTAGCATGGAAGAAACAAAGCAATTTACATAGAGTTTAAATG 2139
361 GATCTCCAGAATC-GAAACTGAAAGAGATTGAATGACTGGCTAACAAAAACAGAGAAAGA 419
2140 GATCTCCAGAATCAGAAACTGAAAGAGTTGAATGACTGGCTAACAAAAACAGAGAAAGA 2199
420 ACAAGGAAATGGAGGAAGAGCCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCAAGTA 479
2200 ACAAGGAAATGGAGGAAGAGCCCTCTTGGACCTGATCTTGAAGACCTTAAACGCCAAGTA 2259
480 CAACAACATAAGGTGCTTCAAGAAGATCTAGAACAAGAAACAAGTCAGGGTCAATTCCTCTC 539
2260 CAACAACATAAGGTGCTTCAAGAAGATCTAGAACAAGAAACAAGTCAGGGTCAATTCCTCTC 2319
540 ACTCACATGGTGGTGGTATGGGAGATCGATGGGCAACACATCTGTAGATGGACAGAACCCGC 599
2320 ACTCACATGGTGGTGGTATGGATGAATCTAGTGGAGATCACGCCAATGCTGCTTTGGAA 2379
600 GAACAACCTTAAGGTATTTGGGAGATCGATGGGCAACACATCTGTAGATGGACAGAACCCGC 659
2380 GAACAACCTTAAGGTATTTGGGAGATCGATGGGCAACACATCTGTAGATGGACAGAACCCGC 2439
660 TGGGTTCTTTTACAAGACATCCTTCTCAAATGGCAACGCTCTTACTGAAGAACAGTGCCTT 719
2440 TGGGTTCTTTTACAAGACATCATAGATTACTGCAACAGTTCCCCCTGGACCTGGAAAAAG 2499
720 TTTAGTGCATGGCTTTCAGAAAAAGAAAGATGCAGTGAACAAGATTTCACACAACCTGGCTTT 779
2500 TTTCTTGCCTGGCTTACAGAAAGCTGAAACAACACTGCCAATGTCTCTACAGGATGCTACCCGT 2559
780 A-----AAGATCAAAATGAAATGTTATCAAGTCTTCAAAAAACTGGCCGTT 824

Db 2560 AAGGAAAGGCTCCTAGAAAGACTCCAAGGGAGTAAAAGAGCTGATGAAACAATGGCAAGAC 2619
QY 825 TTAAAAGCGGATCTAGAAAAGAAAAGCAATCCATGGGCAAACTGTATTCACTCAAAACA 884
Db 2620 CTCCAAGGTGAAATTGAAGCTCACACAGATGTTTATCACAACTGGATGAAAACAGCCAA 2679
QY 885 GATCTTCTTTCAACACTGAAGATAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAG 944
Db 2680 AAAATCCTGAGATCCCTGGAGGTCCGATGATGAGTCCGATGATGAGTCCGATGATGAGT 2739
QY 945 AACTTTGCCCGGTGTTGGGATAATTAGTCCAAAAAACTTGAAAAAGAGTACAGCAC----- 999
Db 2740 AACATGAACCTTCAAGTGGAGTGAACCTTCGGAAAAAAGTCTCTCAACATTAGTCCCATTTG 2799
QY 1000 ----- 999
Db 2800 GAAGCCAGTTCTGACCAGTGGAAACCGTCTGACACCTTTCTCTGCAGGAACCTTCTGGTGTGG 2859
QY 1000 ----- 999
Db 2860 CTACAGCTGAAAGATGAATGAATTAAGCCCGCAGGCACCTATTGGAGGGGACCTTTCCAGCA 2919
QY 1000 ----- 999
Db 2920 GTTCAGAAAGCAGAACGATGTACATAGGGCCCTTCAAGAGGGAATTTGAAAACTAAAGAACCT 2979
QY 1000 ----- 999
Db 2980 GTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTGACAGAGCAGCCCTTTGGAAAGGA 3039
QY 1000 ----- 999
Db 3040 CTAGAGAAACTCTACCAAGAGCCCAAGAGAGCTGCCTCCTGAGGAGAGAGCCCAAGATGTC 3099
QY 1000 ----- 999
Db 3100 ACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACTGAGTGGGAAAAATTTGAACCTG 3159
QY 1000 -----AGACCCCTTGAAAGACTCCAGGAACCTCAA 1028
Db 3160 CACTCCGCTGACTGSCAGAGAAAAAATAGATGAGACCCCTTGAAAGACTCCAGGAACCTCAA 3219
QY 1029 GAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGATCCTGG 1088
Db 3220 GAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCCAAGCTGAGGTGATCAAGGGATCCTGG 3279
QY 1089 CAGCCCGTGGCGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGAGAAAAGTCAAGGCA 1148
Db 3280 CAGCCCGTGGCGATCTCCTCATTTGACTCTCTCCAAGATCACCTCGAGAAAAGTCAAGGCA 3339
QY 1149 CTTGAGGAGAAAATTCGCGCTCTGAAAAGAGAACGTCGAGCCACGTCATGACCTTGCTCGC 1208
Db 3340 CTTGAGGAGAAAATTCGCGCTCTGAAAAGAGAACGTCGAGCCACGTCATGACCTTGCTCGC 3399
QY 1209 CAGCTTACCACCTTTGGGCATTCAGCTCTCACCCGTATAACCTCAGCACCTCTGGAAGACCTG 1268
Db 3400 CAGCTTACCACCTTTGGGCATTCAGCTCTCACCCGTATAACCTCAGCACCTCTGGAAGACCTG 3459
QY 1269 AACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGAA 1328
Db 3460 AACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTGCATGAA 3519
QY 1329 GCCCAGAGGACTTTGGTCCAGCATCTCAGCACCTTCTTTCCACGCTGTGTCAGGGTCCC 1388
Db 3520 GCCCAGAGGACTTTGGTCCAGCATCTCAGCACCTTCTTTCCACGCTGTGTCAGGGTCCC 3579
QY 1389 TGGGAGAGAGCCATCTCGCCAAAACAAAGTGCCCTACTATATCAACACAGAGACTCAAAACA 1448
Db 3580 TGGGAGAGAGCCATCTCGCCAAAACAAAGTGCCCTACTATATCAACACAGAGACTCAAAACA 3639
QY 1449 ACTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGTACCTGAATAAT 1508

Db 1720 CTGCACCTTTCTCTGCAGGAACCTTCTGGTGTGGCTACAGCTGAAAGATGATGAATTAAGC 1779

QY 759 AAGATTACACAACTGGCTTTAAAGATCAAAATGAAATGTTATCAAGTCTTCAAAAACAGT 818

Db 1780 CGGCAGGCACCTATTGGAGCGGACTTTCAGAGAGTTCAGAAAGCAGAACGATGTACATAGG 1839

QY 819 GCCGTTTAAAGCGGATCTAGAAAAGAAAAGCAATCCATGGGCAAACTGTA----- 871

Db 1840 GCCTCAAGAGGGGAATTGAAAACCTAAGAACTAAAGAACTCTAATCATGAGTACTCTTGAGACTGTA 1899

QY 872 -----TCACTCAAAACAAGATCTTCTTTCAACACTGAAGAATAAGTCAGTGACCCAGA 924

Db 1900 CGAATAATTCTGACAGAGCGCCTTTGGAAGGACTAGAGAAAACCTTACCAGGAGCCCGAGA 1959

QY 925 AGACGGAAGCATGGCTGGA-----TAACTTTGCCCGGT 957

Db 1960 GAGCTGCCTCCTGAGGAGAGAGGCCAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTGAG 2019

QY 958 GTTGGGATAATTAGTCCAAAACCTTGAA-----AAGAGTACA 995

Db 2020 GAGGTCAATACTGAGTGGGAAAAAATTGAACCTGCACTCCGCTGACTGGCAGAGAAAAATA 2079

QY 996 GCACAGACCCCTTGAAAGACTCCAGGAACCTCAAGAGACCTTCAAGAGGCCACGGATGAGCTGGACCTCAAG 1055

Db 2080 GATGAGACCCCTTGAAAGACTCCAGGAACCTTCAAGAGGCCACGGATGAGCTGGACCTCAAG 2139

QY 1056 CTGCGCCAAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGGATCTCCTCATTTGAC 1115

Db 2140 CTGCGCCAAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGGATCTCCTCATTTGAC 2199

QY 1116 TCTCTCCAAGATCACCTCGAGAAAGTCAAGGCACTTCGAGGAGAAATTGGCGCTCTGAAA 1175

Db 2200 TCTCTCCAAGATCACCTCGAGAAAGTCAAGGCACTTCGAGGAGAAATTGGCGCTCTGAAA 2259

QY 1176 GAGAACGTGAGCCACGTCAATGACCTTGCTCGCCAGCTTACCACCTTTGGGCAATTCAGCTC 1235

Db 2260 GAGAACGTGAGCCACGTCAATGACCTTGCTCGCCAGCTTACCACCTTTGGGCAATTCAGCTC 2319

QY 1236 TCACCGTATAACCTCAGCACTCTGGAAGACCTGAACACCCAGATGGAAGCTTCTGCAGGTG 1295

Db 2320 TCACCGTATAACCTCAGCACTCTGGAAGACCTGAACACCCAGATGGAAGCTTCTGCAGGTG 2379

QY 1296 GCCGTGAGGACCGAGTCAGGCAGCTGCATGAAGCCACAGGGACTTTGGTCCAGCATCT 1355

Db 2380 GCCGTGAGGACCGAGTCAGGCAGCTGCATGAAGCCACAGGGACTTTGGTCCAGCATCT 2439

QY 1356 CAGCACTTTCTTCCACGTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAA 1415

Db 2440 CAGCACTTTCTTCCACGTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAA 2499

QY 1416 GTGCCCTACTATATCAACCCAGAGACTCAAAACAACTTGCTGGGACCATCCCCAAAATGACA 1475

Db 2500 GTGCCCTACTATATCAACCCAGAGACTCAAAACAACTTGCTGGGACCATCCCCAAAATGACA 2559

QY 1476 GAGCTCTACCACTCTTTAGCTGACCTGAATAATGTCAGATTTCTCAGCTTATAGGACTGCC 1535

Db 2560 GAGCTCTACCACTCTTTAGCTGACCTGAATAATGTCAGATTTCTCAGCTTATAGGACTGCC 2619

QY 1536 ATGAACTCCGAAGACTGCAGAAAGCCCTTTGCTGGATCTCTTGAGCCTGTGAGCTGCA 1595

Db 2620 ATGAACTCCGAAGACTGCAGAAAGCCCTTTGCTGGATCTCTTGAGCCTGTGAGCTGCA 2679

QY 1596 TGTGATGCTTGGACCAGCAGCAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCGAG 1655

Db 2680 TGTGATGCTTGGACCAGCAGCAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCGAG 2739

QY 1656 ATTATTAATTGTTGACCACCTATTTATGACCGCTGGAGCAAGAGCACAACAATTTGGTC 1715

Db 2740 ATTATTAATTGTTGACCACCTATTTATGACCGCTGGAGCAAGAGCACAACAATTTGGTC 2799

QY 1716 AAGTCCCTCTCTGCGTGGATATGTGCTGAAGTGGCTGCTGAATGTTTATGATACGGGA 1775

Db 2800 AACGTCCCTCTCTGCGTGGATATGTGCTGAAGTGGCTGCTGAATGTTTATGATACGGGA 2859

RESULT 11

AAD37258

ID AAD37258 standard; DNA; 4498 BP.

XX

AC AAD37258;

XX

DT 21-AUG-2002 (first entry)

XX

DE Adeno-associated virus vector plasmid, AAV-MCK-3531.

XX

KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;

KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;

KW Becker muscular dystrophy; ds.

XX

OS Chimeric - Homo sapiens.

OS Chimeric - Unidentified.

XX

PN WO200183695-A2.

XX

PD 08-NOV-2001.

XX

PF 27-APR-2001; 2001WO-US13677.

XX

PR 28-APR-2000; 2000US-200777P.

XX

PA (XIAO/) XIAO X.

XX

PI Xiao X;

XX

DR WPI; 2002-049342/06.

XX

PT New dystrophin minigene for treating Duchenne or Becker muscular

PT dystrophy comprises an N-terminal domain or modified N-terminal domain,

PT rod repeats, H1 and H4 domains and a cysteine rich domain of a

PT dystrophin gene -

XX

PS Example 1; Page 62-63; 71pp; English.

XX

CC The present invention relates to an isolated nucleotide sequence encoding

CC a dystrophin minigene. The minigene comprises N-terminal or modified

CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4

CC domains and cysteine-rich domains of dystrophin or utrophin genes. The

CC invention also relates to a recombinant adeno-associated virus (AAV)

CC comprising dystrophin minigene operably linked to an expression control

CC element. The dystrophin minigene in operable linkage with an expression

CC control element, in a recombinant adeno-associated virus or retrovirus is

CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular

CC dystrophy (BMD) in a mammalian subject. The present sequence is AAV

CC vector plasmid construct containing human dystrophin minigenes, a muscle

CC creatine kinase (MCK) promoter and a small polyA signal sequence.

XX

SQ Sequence 4498 BP; 1251 A; 1118 C; 1123 G; 1006 T; 0 other;

Query Match 62.5%; Score 1251.2; DB 24; Length 4498;

Best Local Similarity 77.8%; Pred. No. 0;

Matches 1623; Conservative 0; Mismatches 378; Indels 85; Gaps 6;

```
QY 1 GGCAGTTTCATTGATGGAGAGTGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAA 60
Db 1757 GGCAGTTTCATTGATGGAGAGTGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAA 1816

QY 61 GTATTATCGTGGCTTCTTCTGCTGAGGACACATTGCAAGCACAAAGGAGAGATTCTTAAT 120
Db 1817 GTATTATCGTGGCTTCTTCTGCTGAGGACACATTGCAAGCACAAAGGAGAGATTCTTAAT 1876

QY 121 GATGTGGAAGTGGTGAAGACCAGTTTCTACTACTCATGAGGGGTACATGATGGATTGACA 180
Db 1877 GATGTGGAAGTGGTGAAGACCAGTTTCTACTACTCATGAGGGGTACATGATGGATTGACA 1936

QY 181 GCCCATCAGGGCCGGGTTGGTAATATTCTACAATTTGGGAAGTAAGCTGATTGGAACAGGA 240
Db 1937 GCCCATCAGGGCCGGGTTGGTAATATTCTACAATTTGGGAAGTAAGCTGATTGGAACAGGA 1996

QY 241 AAATTATCAGAAGATGAAGAAACTGAAGTACAAGACGAGATGAATCTCCTAAATTCAAGA 300
Db 1997 AAATTATCAGAAGATGAAGAAACTGAAGTACAAGACGAGATGAATCTCCTAAATTCAAGA 2056

QY 301 TGGGAATGCCTCAGGGTAGCTAGCATGGAAGAACTTAGAAGTGAATTTACATAGAGTTTAAATG 360
Db 2057 TGGGAATGCCTCAGGGTAGCTAGCATGGAAGAACTTAGAAGTGAATTTACATAGAACTCATAGA 2116

QY 361 GATCTCCAGAA-----TCGAAACTGAAGAGTGAATGACTGGCTAAGACCTTGAAGACCTAAAACGC 413
Db 2117 TTAGTGCAACAGTTCCCCCTGGACCTGGAAAAAGTTTCTTGCCCTGGCTTACAGAAAGCTGAA 2176

QY 414 GAAAGAAACAAGGAAAAATGGAGGAAGAGCCCTCTTGGACCTGATCTTGAAGACCTAAAACGC 473
Db 2177 ACAACTGCCAATGTCTACAGGATGCTACCCGTAAAGGAAAGGCTCCTAGAAAGACTCCAAG 2236

QY 474 CAAGTACAACAACATAAGGTCTTCAAGAAGATCTAGAACAAGAACAAAGTCAAGGTCAAT 533
Db 2237 GGAGTAAAAGAGCTGATGAACAATGSCAAGACCTCCAAGGTGAATTTGAAGCTCACACA 2296

QY 534 TCTCTCACTCACATGGTGGTGGTAGTTGATGA-----ATCTAGTGGAGATCAC 581
Db 2297 GATGTTTATCACAACTGGATGAAAACAGCCCAAAAAATCCTGAGATCCCTGGAAAGGTTCC 2356

QY 582 GCAACTGCTGCTTTGGAAGAACAACTTAAGG---TATTGGGAGATCGATGGGCAAAACATC 638
Db 2357 GATGATGCAGTCCCTGTTACAAAGACGTTTGGATAACATGAACCTTCAAGTGGAGTGAACCTT 2416

QY 639 TGTAGATGGACAGAACGCGCTGGGTCTTTTACAAGACATCCTTCTCAAAATGGCAACGT 698
Db 2417 CGGAAAAAGTCTCTCAACATTAGGTCCCATTTGGAAGCCAGTTCTGACCCAGTGGAAAGCGT 2476

QY 699 CTTACTGAAGAACAGTGCCTTTTATGTCATGGCTTTCAGAAAAAAGAAAGATGCAGTGAAC 758
Db 2477 CTGCACCTTCTCTGCAGGAACCTTCTGCTGGCTACAGCTGAAAGATGATGAATTAAGC 2536

QY 759 AAGATTACACAACTGGCTTTAAAGATCAAAAATGAAATGTTATCAAGTCTTCAAAAACTG 818
Db 2537 CGGCAGGCACCTATTGGAGGCGGACTTCCAGCAGTTCAGAAGCAGAACGATGTACATAGG 2596

QY 819 GCCGTTTAAAGCGGATCTAGAAAAAGAAAAAGCAATCCATGGGCAAACTGTA-----871
Db 2597 GCCTTCAAGAGGGAATTGAAAACTTAAAGAACCTGTAATCATGAGTACTCTTGAGACTGTA 2656

QY 872 -----TTCACTCAAACAAGATCTTCTTTCAACACTGAAGAATAAGTCAAGTACCCAGGA 924
Db 2657 CGAATATTCTGACAGAGCAGCCCTTTGGAAGGACTAGAGAACTCTACCGAGAGCCCGAGA 2716

QY 925 AGACGGAAGCATGGCTGGA-----TAACCTTGCCCGGT 957
Db 2717 GAGCTGCCTCCTGAGGAGAGAGGCCAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTGAG 2776

QY 958 GTTGGGATAATTTAGTCCAAAACTTGAA-----AAGAGTACA 995
Db 2777 GAGGTCAATACTAGTGGGAAAAATTTGAACCTGCACTCCGCTGACTGGCAGAGAAAAATA 2836
```

RESULT 12
AAD37230
ID AAD37230 standard; DNA; 4182 BP.

```
QY 996 GCACAGACCCTTTGAAAGACTCCAGGAACCTTCAAGAGGCCACGGATGAGCTGGACCTCAAG 1055
Db 2837 GATGAGACCCTTTGAAAGACTCCAGGAACCTTCAAGAGGCCACGGATGAGCTGGACCTCAAG 2896

QY 1056 CTGCGCCCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGGATCTCCTCATGAC 1115
Db 2897 CTGCGCCCAAGCTGAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGGATCTCCTCATGAC 2956

QY 1116 TCTCTCCAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAATTCGCCCTCTGAAA 1175
Db 2957 TCTCTCCAAGATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAATTCGCCCTCTGAAA 3016

QY 1176 GAGAAACGTGAGCCACGTCATGACCTTGCTCGCCAGCTTACCACCTTTGGCATTCAGCTC 1235
Db 3017 GAGAAACGTGAGCCACGTCATGACCTTGCTCGCCAGCTTACCACCTTTGGCATTCAGCTC 3076

QY 1236 TCACCGTATAACCTCAGCACCTCTGGAAGACCTGAACACACAGATGGAAGCTTCTGCAGGTG 1295
Db 3077 TCACCGTATAACCTCAGCACCTCTGGAAGACCTGAACACACAGATGGAAGCTTCTGCAGGTG 3136

QY 1296 GCCGTCGAGGACCGAGTCAGGCAGCTGTCATGAAGCCCAACAGGGACCTTTGGTCCAGCATCT 1355
Db 3137 GCCGTCGAGGACCGAGTCAGGCAGCTGTCATGAAGCCCAACAGGGACCTTTGGTCCAGCATCT 3196

QY 1356 CAGCACCTTCTTCCACGCTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCCAACAAA 1415
Db 3197 CAGCACCTTCTTCCACGCTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCCAACAAA 3256

QY 1416 GTGCCCTACTATATCAACCACGAGACTCAAAACAACCTTGCTGGACCATCCCAAAATGACA 1475
Db 3257 GTGCCCTACTATATCAACCACGAGACTCAAAACAACCTTGCTGGACCATCCCAAAATGACA 3316

QY 1476 GAGCTCTACCAGTCTTTAGCTGACCTGAATAATGTCAAGATCTCTGAGCCTTATAGGACTGCC 1535
Db 3317 GAGCTCTACCAGTCTTTAGCTGACCTGAATAATGTCAAGATCTCTGAGCCTTATAGGACTGCC 3376

QY 1536 ATGAAACTCCGAAGACTGCAGAAGGCCCTTTGCTTGGATCTCTTGAAGCCTGTAGCTGCA 1595
Db 3377 ATGAAACTCCGAAGACTGCAGAAGGCCCTTTGCTTGGATCTCTTGAAGCCTGTAGCTGCA 3436

QY 1596 TGTGATGCCTTTGGACCAGCACACCTCAAGCAAAAATGACCAGCCCATGGATATCCTGCAG 1655
Db 3437 TGTGATGCCTTTGGACCAGCACACCTCAAGCAAAAATGACCAGCCCATGGATATCCTGCAG 3496

QY 1656 ATTATTAAATTGTTTGACCACCTATTATGACCCGCTTGTGAGCAAGAGCACACAATTTGGTC 1715
Db 3497 ATTATTAAATTGTTTGACCACCTATTATGACCCGCTTGTGAGCAAGAGCACACAATTTGGTC 3556

QY 1716 AACGTCCTCTCTGCGTGGATATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGA 1775
Db 3557 AACGTCCTCTCTGCGTGGATATGTGTCTGAACCTGGCTGCTGAATGTTTATGATACGGGA 3616

QY 1776 CGAACAGGGAGGATCCGTGTCTGTCTTTTAAAACTGGCATCATTTCCCTGTGTAAGCA 1835
Db 3617 CGAACAGGGAGGATCCGTGTCTGTCTTTTAAAACTGGCATCATTTCCCTGTGTAAGCA 3676

QY 1836 CATTTGGAAGACAAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTGT 1895
Db 3677 CATTTGGAAGACAAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTGT 3736

QY 1896 GACCAGCGCAGGCTGGGCCCTCCTTCTGTCATGATTTCTATCCAAATTTCCAAGACAGTTGGGT 1955
Db 3737 GACCAGCGCAGGCTGGGCCCTCCTTCTGTCATGATTTCTATCCAAATTTCCAAGACAGTTGGGT 3796

QY 1956 GAAGTTGCATCCTTTGGGGGCGAGTAACATTGAGCCAAAGTGTCCGGA 2001
Db 3797 GAAGTTGCATCCTTTGGGGGCGAGTAACATTGAGCCAAAGTGTCCGGA 3842
```


QY 1000 ----- 999
Db 2380 TCTCTGCAGGAACCTTCTGGTGTGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCA 2439
QY 1000 ----- 999
Db 2440 CCTATTGGAGGCGACTTTCAGCAGTTCAGAAAGCAGAACGATGTACATAGGCGCTTCAAG 2499
QY 1000 ----- 999
Db 2500 AGGGAATTGAAAACCTAAAGAACCTGTAAATCATGAGTACTCTTGAGACTGTACGAATATT 2559
QY 1000 ----- 999
Db 2560 CTGACAGAGCAGCCTTTTGGAAAGGACTAGAGAAACTCTACCAGAGCCCCAGAGAGCTGCCT 2619
QY 1000 ----- 999
Db 2620 CCTGAGGAGAGAGCCCAAGATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAAT 2679
QY 1000 -----AGACC 1004
Db 2680 ACTGAGTGGGAAAAATTGAACCTGCACCTCCGCTGACTGGCAGAGAAAAATAGATGAGACC 2739
QY 1005 CTTGAAAGACTCCAGGAACTTCAAGAGGCGCACGGATGAGCTGGAGCTCAAGCTCGGCCAA 1064
Db 2740 CTTGAAAGACTCCAGGAACTTCAAGAGGCGCACGGATGAGCTGGAGCTCAAGCTCGGCCAA 2799
QY 1065 GCTGAGGTGATCAAGGGATCCTGGCAGGCCCGTGGCGGATCTCCTCATTTGACTCTCTCCAA 1124
Db 2800 GCTGAGGTGATCAAGGGATCCTGGCAGGCCCGTGGCGGATCTCCTCATTTGACTCTCTCCAA 2859
QY 1125 GATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAATTGGCGCTCTGAAAGAGAACGTG 1184
Db 2860 GATCACCTCGAGAAAGTCAAGGCACCTTCGAGGAGAAATTGGCGCTCTGAAAGAGAACGTG 2919
QY 1185 AGCCACGTCATGACCTTGCTCGCCAGCTTACCACCTTTGGGCATTCAGCTCTCACCGTAT 1244
Db 2920 AGCCACGTCATGACCTTGCTCGCCAGCTTACCACCTTTGGGCATTCAGCTCTCACCGTAT 2979
QY 1245 AACCTCAGCACTCTGGAAGACCTGAACACCCAGATGGAAGCTTTCGCCAGGCTGCGCGTCGAG 1304
Db 2980 AACCTCAGCACTCTGGAAGACCTGAACACCCAGATGGAAGCTTTCGCCAGGCTGCGCGTCGAG 3039
QY 1305 GACCGAGTCAGGCAGCTGCATGAAGCCCAAGCCAGAGGACTTTGGTCCAGCATCTCAGCACTTT 1364
Db 3040 GACCGAGTCAGGCAGCTGCATGAAGCCCAAGCCAGAGGACTTTGGTCCAGCATCTCAGCACTTT 3099
QY 1365 CTTTCCACGTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTAC 1424
Db 3100 CTTTCCACGTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTAC 3159
QY 1425 TATATCAACCACGAGACTCAAAACAACCTTGCTGGGACCATCCCAAAATGACAGAGCTCTAC 1484
Db 3160 TATATCAACCACGAGACTCAAAACAACCTTGCTGGGACCATCCCAAAATGACAGAGCTCTAC 3219
QY 1485 CAGTCTTTAGCTGACCTGAATAATGTGATGATCTCAGCTTATAGGACTGCCATGAAACTC 1544
Db 3220 CAGTCTTTAGCTGACCTGAATAATGTGATGATCTCAGCTTATAGGACTGCCATGAAACTC 3279
QY 1545 CGAAGACTGCAGAGGCGCTTTGCTTGGATCTCTTGGCCCTGTGAGCTGATGATGCC 1604
Db 3280 CGAAGACTGCAGAGGCGCTTTGCTTGGATCTCTTGGCCCTGTGAGCTGATGATGCC 3339
QY 1605 TTGGACCAGCACAACTCAAGCAAAAATGACCAAGCCCATGGATATCCTGCAGATTATTAAT 1664
Db 3340 TTGGACCAGCACAACTCAAGCAAAAATGACCAAGCCCATGGATATCCTGCAGATTATTAAT 3399
QY 1665 TGTGTTGACCACTATTATGACCGCCTGGAGCAAGAGCACAACAATTTGGTCAACGTCCT 1724
Db 3400 TGTGTTGACCACTATTATGACCGCCTGGAGCAAGAGCACAACAATTTGGTCAACGTCCT 3459
QY 1725 CTCTGCGTGGATATGTCTGAACTGGCTGCTGAAATGTTTATGATACGGGACGACAGG 1784

Db 3460 CTCTGCGTGGATATGTCTGAACTGGCTGCTGAATGTTATGATACGGGACGACAGGG 3519
QY 1785 AGGATCCGTCCTGTCTTTTAAAACTGGCATCATTTCCCTGTGTAAGCACATTTGGAA 1844
Db 3520 AGGATCCGTCCTGTCTTTTAAAACTGGCATCATTTCCCTGTGTAAGCACATTTGGAA 3579
QY 1845 GACAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAGCGC 1904
Db 3580 GACAAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAGCGC 3639
QY 1905 AGGCTGGGCTCCTTCTGTCATGATTTCTATCCAAATTTCCAAGACAGTTGGGTGAAGTTGCA 1964
Db 3640 AGGCTGGGCTCCTTCTGTCATGATTTCTATCCAAATTTCCAAGACAGTTGGGTGAAGTTGCA 3699
QY 1965 TCCTTTGGGGCAGTAACATTTAGCCAAAGTGTCGGGA 2001
Db 3700 TCCTTTGGGGCAGTAACATTTAGCCAAAGTGTCGGGA 3736

RESULT 13

AAD37234
ID AAD37234 standard; DNA; 3999 BP.
XX
AC AAD37234;
XX
DT 21-AUG-2002 (first entry)
XX
DE Human dystrophin minigene delta3990.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Homo sapiens.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-US13677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
XX
PI xiao X;
XX
DR WPI; 2002-049342/06.
XX
PT New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -
XX
PS Example 1; Page 46-47; 71pp; English.
XX
CC The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The
CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is human
CC dystrophin minigene delta3990 containing nucleotides 1-1668 (N-terminus,
CC hinge H1 and rods R1 and R2), 8059-10227 (rods R22, R23 and R24, hinge
CC H4 and CR domain) and 11047-11058 (dystrophin last 3 amino acids).
XX
SQ Sequence 3999 BP; 1223 A; 907 C; 933 G; 936 T; 0 other;

Query Match				55.6%;	Score 1112;	DB 24;	Length 3999;
Best Local Similarity				71.1%;	Pred. No. 1.9e-312;		
Matches 1816;				Conservative 0;	Mismatches 185;	Indels 553;	Gaps 4;
1	GGCAGTTCATTGATGGAGAGTGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAGAA	60					
1000	GGCAGTTCATTGATGGAGAGTGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAGAA	1059					
61	GTATTATCGTGGCTTCTTCTGCTGAGGACACATTTCAAGCACAAAGGAGAGATTTCCTAAT	120					
1060	GTATTATCGTGGCTTCTTCTGCTGAGGACACATTTCAAGCACAAAGGAGAGATTTCCTAAT	1119					
121	GATGTGGAAGTGGTGAAGACCAAGTTTCATACTCATGAGGGGTACATGATGGATTGACA	180					
1120	GATGTGGAAGTGGTGAAGACCAAGTTTCATACTCATGAGGGGTACATGATGGATTGACA	1179					
181	GCCCATCAGGCGGGTGGTAAATATTTACAATTGGGAAGTAAGCTGATTGGAACAGGA	240					
1180	GCCCATCAGGCGGGTGGTAAATATTTACAATTGGGAAGTAAGCTGATTGGAACAGGA	1239					
241	AAATTATCAGAAAGATGAAGAAACTGAAGTACAAGACGACAGATGAATCTCCTAAATTCAAGA	300					
1240	AAATTATCAGAAAGATGAAGAAACTGAAGTACAAGACGACAGATGAATCTCCTAAATTCAAGA	1299					
301	TGGGAATGCCTCAGGGTAGCTAGCATGGAAAAACAAGCAATTTACATAGAGTTTTAATG	360					
1300	TGGGAATGCCTCAGGGTAGCTAGCATGGAAAAACAAGCAATTTACATAGAGTTTTAATG	1359					
361	GATCTCCAGAATC-GAAACTGAAAGAGTTGAATGACTGGCTAAACAAAAACAGAAAGA	419					
1360	GATCTCCAGAATCAGAAACTGAAAGAGTTGAATGACTGGCTAAACAAAAACAGAAAGA	1419					
420	ACAAGGAAAAATGGAGGAAGAGCCTCTTGGACCTGATCTTGAAGACCTAAACGCCAAGTA	479					
1420	ACAAGGAAAAATGGAGGAAGAGCCTCTTGGACCTGATCTTGAAGACCTAAACGCCAAGTA	1479					
480	CAACAACATAAGGTGCTTCAAGAAGATCTAGAACAAAGAACAAAGTCAGGGTCAATTCTCTC	539					
1480	CAACAACATAAGGTGCTTCAAGAAGATCTAGAACAAAGAACAAAGTCAGGGTCAATTCTCTC	1539					
540	ACTCACATGGTGGTGGTAGTTGATGAATCTAGTGGAGATCACGCAACTGCTTTGGAA	599					
1540	ACTCACATGGTGGTGGTAGTTGATGAATCTAGTGGAGATCACGCAACTGCTTTGGAA	1599					
600	GAACAACCTTAAGGTATTGGGAGATCGATGGGCAACATCTGTAGATGGACAGAACCCGC	659					
1600	GAACAACCTTAAGGTATTGGGAGATCGATGGGCAACATCTGTAGATGGACAGAACCCGC	1659					
660	TGGGTTCTTTTACAAGAC-----	677					
1660	TGGGTTCTTTTACAAGACCCAGCCTGACCTAGCTCCTGGACTGACCACTATTGGAGCCTCT	1719					
678	-----	677					
1720	CCTACTCAGACTGTTACTCTGGTGACACAACCTGTGGTTACTAAGGAAACTGCCATCTCC	1779					
678	-----ATCCTTCTCAAAATGGCAACGT	698					
1780	AAACTAGAAATGCCATCTTCCCTTGATGTTGGAGGTACCTACTCATAGATTACTGCAACAG	1839					
699	CTTACTGAAGAACAGTGCCTTTTATAGTCATGGCTTTCAGAAAAAAGAGATGCAGTGAAC	758					
1840	TTCCCCCTGGACCTGGAAAAGTTCTTGCCTGGCTTACAGAAGCTGAAACAACACTGCCAAT	1899					
759	AAGATTACACAACTGGCTTTA-----AAGATCAAAATGAAATGTTATCA	803					
1900	GTCCTACAGGATGCTACCCGTAAAGGAAAGGCTCCTAGAAGACTCCAAGGGAGTAAAGAG	1959					
804	AGTCTTCAAAAACTGGCCGTTTAAAGCGGATCTAGAAAAAGAAAAAGCAATCCATGGGC	863					
1960	CTGATGAACAATGGCAAGACCTCCAAGGTGAAATTTGAAGCTCACACAGATGTTTATCAC	2019					
864	AAACTGTATTCACCTCAACAAGATCTTCTTTCAACACTGAAGAATAAGTCAGTGACCCAG	923					

Db	2020	AACCTGGATGAAAAACAGCCCAAAAAATCCTGAGATCCCTGGAAAGTTCCGATGATGCAGTC	2079
QY	924	AAGACGGAAGCATGGCTGGATAAACTTTGCCCGGTGTTGGGATAAATTTAGTCCAAAAACTT	983
Db	2080	CTGTTACAAAGACGTTTGGATAACATGAACCTTCAAGTGGAGTGAACCTTCGGAAGAAAGTCT	2139
QY	984	GAAAAGAGTACAGCAC-----	999
Db	2140	CTCAACATTAGGTCCCATTTGGAAGCCAGTTCTGACCAAGTGAAGCGTCTGCACCTTTCT	2199
QY	1000	-----	999
Db	2200	CTGCAGGAACCTTCTGGTGTGGCTACAGCTGAAAGATGATGAATTAAGCCGCGCAGGCACCT	2259
QY	1000	-----	999
Db	2260	ATTGGAGCGACTTCCAGCAGTTTCAAGAGCAGAACGATGTACATAGGGCCTTCAAGAGG	2319
QY	1000	-----	999
Db	2320	GAATTGAAAACTAAAGAAACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTTCTG	2379
QY	1000	-----	999
Db	2380	ACAGAGCAGCCTTTTGAAGGACTAGAGAAAACTCTACCAGSAGCCCAGAGAGCTGCCTCCT	2439
QY	1000	-----	999
Db	2440	GAGGAGAGAGCCCAAGAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACT	2499
QY	1000	-----AGACCCCTT	1007
Db	2500	GAGTGGAAAAAATTGAACCTGCACCTCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTT	2559
QY	1008	GAAAGACTCCAGGAACCTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCAAAGCT	1067
Db	2560	GAAAGACTCCAGGAACCTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTGCGCCAAAGCT	2619
QY	1068	GAGTGATCAAGGGATCCTGGCAGCCCGTGGCGATCTCCTCATTTGACTCTCTCCAAGAT	1127
Db	2620	GAGTGATCAAGGGATCCTGGCAGCCCGTGGCGATCTCCTCATTTGACTCTCTCCAAGAT	2679
QY	1128	CACCTCGAGAAAAGTCAAGGCACCTTCGAGGAGAAAATTCGCCCTCTGAAAGAGAACGTGAGC	1187
Db	2680	CACCTCGAGAAAAGTCAAGGCACCTTCGAGGAGAAAATTCGCCCTCTGAAAGAGAACGTGAGC	2739
QY	1188	CAGTCAATGACCTTGCTCGCCAGCTTACCACCTTTGGGCATTCAGCTCTCACCCGTATAAC	1247
Db	2740	CAGTCAATGACCTTGCTCGCCAGCTTACCACCTTTGGGCATTCAGCTCTCACCCGTATAAC	2799
QY	1248	CTCAGCACTCTGGAAGACCTGAACACACAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGAC	1307
Db	2800	CTCAGCACTCTGGAAGACCTGAACACACAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGAC	2859
QY	1308	CGAGTCAGGCAGCTGCATGAAGCCCAACAGGGACTTTGGTCCAGCATCTCAGCACTTCTT	1367
Db	2860	CGAGTCAGGCAGCTGCATGAAGCCCAACAGGGACTTTGGTCCAGCATCTCAGCACTTCTT	2919
QY	1368	TCCACGTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTAT	1427
Db	2920	TCCACGTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCCTACTAT	2979
QY	1428	ATCAACCCAGAGACTCAAAACAACTTGTGGGACCATCCCAAAATGACAGAGCTCTACCAG	1487
Db	2980	ATCAACCCAGAGACTCAAAACAACTTGTGGGACCATCCCAAAATGACAGAGCTCTACCAG	3039
QY	1488	TCTTTAGCTGACCTGAATAATGTGAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGA	1547
Db	3040	TCTTTAGCTGACCTGAATAATGTGAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGA	3099
QY	1548	AGACTGCAGAAAGGCCCTTTGCTTGGATCTCTTGGCCCTGTGAGCTGTGATGCCTTG	1607

Db 3100 AGACTGCAGAAAGGCCCTTTGGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTG 3159
QY 1608 GACCAGCACAAACCTCAAGCAAAATGACCAGCCCATGGATATCCTCGACATTAATTTGT 1667
Db 3160 GACCAGCACAAACCTCAAGCAAAATGACCAGCCCATGGATATCCTCGACATTAATTTGT 3219
QY 1668 TTGACCACTATTATGACCGCTGGAGCAAGAGCACAACAATTTGGTCAACGTCCCTCTC 1727
Db 3220 TTGACCACTATTATGACCGCTGGAGCAAGAGCACAACAATTTGGTCAACGTCCCTCTC 3279
QY 1728 TGCCTGGATATGTCTGAACCTGGCTGGCTGAATGTTATGATACGGGACAAACAGGGAGG 1787
Db 3280 TGCCTGGATATGTCTGAACCTGGCTGGCTGAATGTTATGATACGGGACAAACAGGGAGG 3339
QY 1788 ATCCGTGCTCTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACAATTTGGAAGAC 1847
Db 3340 ATCCGTGCTCTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACAATTTGGAAGAC 3399
QY 1848 AAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCACAGGATTTTGTGACCGCAGG 1907
Db 3400 AAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCACAGGATTTTGTGACCGCAGG 3459
QY 1908 CTGGGCTCTCTCTGTCATGATTTCTATCCAAATTCACAGACAGTTGGGTGAAGTTCATCC 1967
Db 3460 CTGGGCTCTCTCTGTCATGATTTCTATCCAAATTCACAGACAGTTGGGTGAAGTTCATCC 3519
QY 1968 TTTGGGGGAGTAACATTTGAGCCAAAGTGTCCCGA 2001
Db 3520 TTTGGGGGAGTAACATTTGAGCCAAAGTGTCCCGA 3553

RESULT 14

AAD37256
ID AAD37256 standard; DNA; 4966 BP.

XX AAD37256;

XX 21-AUG-2002 (first entry)

DE Adeno-associated virus vector plasmid, AAV-MCK-delta3990.

XX Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.

XX Chimeric - Homo sapiens.
OS Chimeric - Unidentified.

XX WO200183695-A2.

PN 08-NOV-2001.

XX 27-APR-2001; 2001WO-US13677.

PR 28-APR-2000; 2000US-200777P.

XX (XIAO/) XIAO X.

PI XIAO X;

XX WPI; 2002-049342/06.

XX New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -

XX Example 1; Page 59-60; 71pp; English.

XX The present invention relates to an isolated nucleotide sequence encoding
CC a dystrophin minigene. The minigene comprises N-terminal or modified
CC N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
CC domains and cysteine-rich domains of dystrophin or utrophin genes. The

CC invention also relates to a recombinant adeno-associated virus (AAV)
CC comprising dystrophin minigene operably linked to an expression control
CC element. The dystrophin minigene in operable linkage with an expression
CC control element, in a recombinant adeno-associated virus or retrovirus is
CC useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
CC dystrophy (BMD) in a mammalian subject. The present sequence is AAV
CC vector plasmid construct containing human dystrophin minigenes, a muscle
CC creatine kinase (MCK) promoter and a small polyA signal sequence.

XX Sequence 4966 BP; 1403 A; 1216 C; 1232 G; 1115 T; 0 other;

QY Query Match 55.6%; Score 1112; DB 24; Length 4966;
Best Local Similarity 71.1%; Pred. No. 2.1e-312;
Matches 1816; Conservative 0; Mismatches 185; Indels 553; Gaps 4;

QY 1 GGCAGTTTCATTTGAGAGTGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAA 60

Db 1757 GGCAGTTTCATTTGAGAGTGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAAGAA 1816

QY 61 GTATTATCGTGGCTTCTTCTGCTGAGGACACATTTGAAGCACAAGGAGAGATTTCTAAT 120

Db 1817 GTATTATCGTGGCTTCTTCTGCTGAGGACACATTTGAAGCACAAGGAGAGATTTCTAAT 1876

QY 121 GATGTGGAAGTGGTGAAGACCAGTTTTCATCTACTCATGAGGGTACATGATGGATTTGACA 180

Db 1877 GATGTGGAAGTGGTGAAGACCAGTTTTCATCTACTCATGAGGGTACATGATGGATTTGACA 1936

QY 181 GCCCATCAGGCGCGGTTGGTAATATTCTACAAATTTGGGAAGTAAGCTGATTGGAACAGGA 240

Db 1937 GCCCATCAGGCGCGGTTGGTAATATTCTACAAATTTGGGAAGTAAGCTGATTGGAACAGGA 1996

QY 241 AAATTATCAGAAGATGAAGAAACTGAAGTACAAGACACAGATGAATCTCTAAATTTCAAGA 300

Db 1997 AAATTATCAGAAGATGAAGAAACTGAAGTACAAGACACAGATGAATCTCTAAATTTCAAGA 2056

QY 301 TGGGAATGCCCTCAGGCTAGCTAGCATGGAAAAACAAGCAATTTACATAGAGTTTAAATG 360

Db 2057 TGGGAATGCCCTCAGGCTAGCTAGCATGGAAAAACAAGCAATTTACATAGAGTTTAAATG 2116

QY 361 GATCTCCAGAATC-GAAACTGAAAGAGTTGAATGATCTGGCTAACAAAAACAGAAAGA 419

Db 2117 GATCTCCAGAATCAGAAACTGAAAGAGTTGAATGATCTGGCTAACAAAAACAGAAAGA 2176

QY 420 ACAAGGAAAAATGGAGAGAGCCTCTTGGACCTGATCTTGAAGACCTAAACGCCAAGTA 479

Db 2177 ACAAGGAAAAATGGAGAGAGCCTCTTGGACCTGATCTTGAAGACCTAAACGCCAAGTA 2236

QY 480 CAACAACATAAGGTGCTTCAAGAAGATCTAGAACAAGCAAGTCAAGGTCAATCTCTC 539

Db 2237 CAACAACATAAGGTGCTTCAAGAAGATCTAGAACAAGCAAGTCAAGGTCAATCTCTC 2296

QY 540 ACTCACATGGTGGTAGTTGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAA 599

Db 2297 ACTCACATGGTGGTAGTTGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAA 2356

QY 600 GAACAACCTTAAGGTATTGGGAGATCGATGGGCAACACATCTGTAGATGGACAGAACCCG 659

Db 2357 GAACAACCTTAAGGTATTGGGAGATCGATGGGCAACACATCTGTAGATGGACAGAACCCG 2416

QY 660 TGGGTTCTTTTACAAGAC----- 677

Db 2417 TGGGTTCTTTTACAAGACCCAGCTGACCTAGCTCCTGGACTGACCACCTATTTGGAGCCTCT 2476

QY 678 ----- 677

Db 2477 CCTACTCAGACTGTTACTCTGTTGACACAACCTGTGGTTACTAAGGAAACTGCCATCTCC 2536

QY 678 -----ATCCTTCTCAAAATGGCAACGT 698

Db 2537 AAAGTAAAAATGCCATCTTCTCTTGAATGTTGGAGGTACCTACTCATAGATTACTGCAACAG 2596

QY 699 CTTACTGAAGAACAGTGCCTTTTGTAGTGCATGGCTTTTCAGAAAAAAGATGCAGTGAAC 758

Db 2597 TTCCCCCTGGACCTGGAAAGTTTCTTGCCCTGGCTTACAGAAAGCTGAAACAACTGCCAAT 2656
QY 759 AAGATTACACAACTGGCTTTA-----AAGATCAAAATGAAATGTTATCA 803
Db 2657 GTCTACAGGATGCTACCCGTAAGGAAAGGCTCTAGAAGACTCCAAGGGAGTAAAGAG 2716
QY 804 AGTCTTCAAAAACCTGGCCGTTTAAAGCGGATCTAGAAAAGAAAAGCAATCCATGGGC 863
Db 2717 CTGATGAAACAATGGCAAGACCTCCAAGGTGAAATGAAAGCTCACACAGATGTTTATCAC 2776
QY 864 AAACGTGATTTCACTCAAAACAAGATCTTCTTCAACACTGAAGAATAAGTCAGTGACCCAG 923
Db 2777 AACCTGGATGAAAACAGCCCAAAAATCCTGAGATCCCTGGAAGGTTCCGATGATGCAGTC 2836
QY 924 AAGACGGAAGCATGGCTGGATAACTTTGCCCGGTGTTGGGATAATTTAGTCCAAAAACCT 983
Db 2837 CTGTTACAAAGACGTTTGGATAACATGAACCTCAAGTGGAGTGAACCTCGGAAAAAGTCT 2896
QY 984 GAAAAGAGTACAGCAC-----999
Db 2897 CTCACATTAGGTCCCATTTGGAAGCCAGTTCTGACCAGTGGAAAGCGTCTGCACCTTTCT 2956
QY 1000 -----999
Db 2957 CTGACGAACTTCTGGTGTGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCT 3016
QY 1000 -----999
Db 3017 ATTGGAGCGGACTTTCAGCAGTTTCAGAAAGCAGAACGATGTACATAGGGCCTTCAAGAGG 3076
QY 1000 -----999
Db 3077 GAATTGAAAACCTAAAGAACCCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTCTG 3136
QY 1000 -----999
Db 3137 ACAGAGCAGCCTTTGGAAGGACTAGAGAAACTCTACCAGGAGCCCAGAGAGCTGCCCTCCT 3196
QY 1000 -----999
Db 3197 GAGGAGAGAGCCCAAGAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACT 3256
QY 1000 -----AGACCCCTT 1007
Db 3257 GAGTGGGAAAAATTGAACCTGCACCTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTT 3316
QY 1008 GAAAGACTCCAGGAACTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTCGGCCAAAGCT 1067
Db 3317 GAAAGACTCCAGGAACTTCAAGAGGCCACGGATGAGCTGGACCTCAAGCTCGGCCAAAGCT 3376
QY 1068 GAGGTGATCAAGGGATCCTGGCAGCCGCTGGCGGATCTCCTCATTTGACTCTCTCCAAAGAT 1127
Db 3377 GAGGTGATCAAGGGATCCTGGCAGCCGCTGGCGGATCTCCTCATTTGACTCTCTCCAAAGAT 3436
QY 1128 CACCTCGAGAAAAGTCAAGGCACTTCGAGGAGAAAATTGGCGCTCTGAAAGAGAACGTTGAGC 1187
Db 3437 CACCTCGAGAAAAGTCAAGGCACTTCGAGGAGAAAATTGGCGCTCTGAAAGAGAACGTTGAGC 3496
QY 1188 CACGTCAATGACCTTGCTCGCCAGCTTACCACCTTTGGGCAATTCAGCTCTCACCGTATAAC 1247
Db 3497 CACGTCAATGACCTTGCTCGCCAGCTTACCACCTTTGGGCAATTCAGCTCTCACCGTATAAC 3556
QY 1248 CTCAGCACTCTGGAAGACCTGAAACACACAGATGGAAGCTTCTGCAGGTGGCGTTCGAGGAC 1307
Db 3557 CTCAGCACTCTGGAAGACCTGAAACACACAGATGGAAGCTTCTGCAGGTGGCGTTCGAGGAC 3616
QY 1308 CGAGTCAGGCAGCTGCTGAAGCCCAACAGGGACTTTGGTCCAGCATCTCAGCACTTTCTT 1367
Db 3617 CGAGTCAGGCAGCTGCTGAAGCCCAACAGGGACTTTGGTCCAGCATCTCAGCACTTTCTT 3676
QY 1368 TCCACGTCTGTCCAGGTCCTTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCTTACTAT 1427
Db 3677 TCCACGTCTGTCCAGGTCCTTGGGAGAGAGCCATCTCGCCAAACAAAGTGCCTTACTAT 3736

QY 1428 ATCAACCACGAGACTCAAAACAACCTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAG 1487
Db 3737 ATCAACCACGAGACTCAAAACAACCTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAG 3796
QY 1488 TCTTTAGCTGACCTGAATAATGTCAAGATTTCTAGCTTATAGGACTGCCATGAAACTCCGA 1547
Db 3797 TCTTTAGCTGACCTGAATAATGTCAAGATTTCTAGCTTATAGGACTGCCATGAAACTCCGA 3856
QY 1548 AGACTGCAGAAAGGCCCTTTGCTTGGATCTCTTGACCTGTGACCTGTGATGATGCCCTT 1607
Db 3857 AGACTGCAGAAAGGCCCTTTGCTTGGATCTCTTGACCTGTGACCTGTGATGATGCCCTT 3916
QY 1608 GACCAGCACAAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGT 1667
Db 3917 GACCAGCACAAACCTCAAGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGT 3976
QY 1668 TTGACCACCTATTATGACCGCCTGGAGCAAGAGCACACAACAATTTGGTCAACGTCCCTCTC 1727
Db 3977 TTGACCACCTATTATGACCGCCTGGAGCAAGAGCACACAACAATTTGGTCAACGTCCCTCTC 4036
QY 1728 TCGGTGGATATGTCTGAACTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGG 1787
Db 4037 TCGGTGGATATGTCTGAACTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGG 4096
QY 1788 ATCCGTGTCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGAC 1847
Db 4097 ATCCGTGTCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGAC 4156
QY 1848 AAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAGCGCAGG 1907
Db 4157 AAGTACAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAGCGCAGG 4216
QY 1908 CTGGGCCCTCCTTCTGTCATGATTCTATCCAAATTTCCAAAGACAGTTGGGTGAAGTTGCATCC 1967
Db 4217 CTGGGCCCTCCTTCTGTCATGATTCTATCCAAATTTCCAAAGACAGTTGGGTGAAGTTGCATCC 4276
QY 1968 TTTGGGGGCAGTAACATTTGAGCCCAAGTTGCCCGGA 2001
Db 4277 TTTGGGGGCAGTAACATTTGAGCCCAAGTTGCCCGGA 4310

RESULT 15

AAD37262
ID AAD37262 standard; DNA; 4990 BP.
XX
AC AAD37262;
XX
DT 21-AUG-2002 (first entry)
XX
DE Adeno-associated virus (AAV) vector plasmid, AAV-CMV-delta3990.
XX
KW Human; dystrophin minigene; muscular; gene therapy; utrophin; spectrin;
KW adeno-associated virus; AAV; Duchenne muscular dystrophy; DMD; BMD;
KW Becker muscular dystrophy; ds.
XX
OS Chimeric - Homo sapiens.
OS Chimeric - Cytomegalovirus.
OS Chimeric - Unidentified.
XX
PN WO200183695-A2.
XX
PD 08-NOV-2001.
XX
PF 27-APR-2001; 2001WO-USI3677.
XX
PR 28-APR-2000; 2000US-200777P.
XX
PA (XIAO/) XIAO X.
XX
PI xiao X;
XX
DR WPI; 2002-049342/06.

XX New dystrophin minigene for treating Duchenne or Becker muscular
PT dystrophy comprises an N-terminal domain or modified N-terminal domain,
PT rod repeats, H1 and H4 domains and a cysteine rich domain of a
PT dystrophin gene -
XX
PS
XX
XX Example 1; Page 67-68; 71pp; English.
The present invention relates to an isolated nucleotide sequence encoding
a dystrophin minigene. The minigene comprises N-terminal or modified
N-terminal domains, rod repeats, utrophin or a spectrin gene, H1 and H4
domains and cysteine-rich domains of dystrophin or utrophin genes. The
invention also relates to a recombinant adeno-associated virus (AAV)
comprising dystrophin minigene operably linked to an expression control
element. The dystrophin minigene in operable linkage with an expression
control element, in a recombinant adeno-associated virus or retrovirus is
useful for treating Duchenne muscular dystrophy (DMD) and Becker muscular
dystrophy (BMD) in a mammalian subject. The present sequence is AAV
vector plasmid construct containing human dystrophin minigenes, a
cytomegalovirus (CMV) promoter and a small polyA signal sequence.
Sequence 4990 BP; 1439 A; 1185 C; 1208 G; 1158 T; 0 other;
Query Match 55.6%; Score 1112; DB 24; Length 4990;
Best Local Similarity 71.1%; Pred. No. 2.le-312;
Matches 1816; Conservative 0; Mismatches 185; Indels 553; Gaps 4;
QY 1 GGCAGTTCATTGATGGAGAGTGAAGTAAACCTGGACCGTTATCAACACAGCTTTAGAAGAA 60
Db 1781 GGCAGTTCATTGATGGAGAGTGAAGTAAACCTGGACCGTTATCAACACAGCTTTAGAAGAA 1840
QY 61 GTATTATCGTGGCTTCTTCTGCTGAGGACACATTGCAAGCACACAGGAGAGATTCTTAAT 120
Db 1841 GTATTATCGTGGCTTCTTCTGCTGAGGACACATTGCAAGCACACAGGAGAGATTCTTAAT 1900
QY 121 GATGTGGAAGTGGTGAAGACAGTTTCATACCTCATGAGGGGTACATGATGGATTGACA 180
Db 1901 GATGTGGAAGTGGTGAAGACAGTTTCATACCTCATGAGGGGTACATGATGGATTGACA 1960
QY 181 GCCCATCAGGGCCGGTTGGTAATATCTACAATTTGGGAAGTAAGCTGATGGAAACAGGA 240
Db 1961 GCCCATCAGGGCCGGTTGGTAATATCTACAATTTGGGAAGTAAGCTGATGGAAACAGGA 2020
QY 241 AAATTATCAGAAGATGAAGAACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGA 300
Db 2021 AAATTATCAGAAGATGAAGAACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCAGA 2080
QY 301 TGGGAATGCCTCAGGGTAGCTAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATG 360
Db 2081 TGGGAATGCCTCAGGGTAGCTAGCATGGAAAAACAAAGCAATTTACATAGAGTTTAAATG 2140
QY 361 GATCTCCAGAAATC-GAAACTGAAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGA 419
Db 2141 GATCTCCAGAAATCAGAAACTGAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGA 2200
QY 420 ACAAGGAAAAATGGAGGAAGAGCCCTTTGGACCTGATCTTGAAGACCTAAACGCCAAGTA 479
Db 2201 ACAAGGAAAAATGGAGGAAGAGCCCTTTGGACCTGATCTTGAAGACCTAAACGCCAAGTA 2260
QY 480 CAACAACATAAAGGTGCTTCAAGAAGATCTAGAAACAAGAACAAGTCAGGGTCAATTTCTCTC 539
Db 2261 CAACAACATAAAGGTGCTTCAAGAAGATCTAGAAACAAGAACAAGTCAGGGTCAATTTCTCTC 2320
QY 540 ACTCACATGGTGGTGGTAGTTGATGAATCTAGTGGAGATCACGCAACTGCTGCTTGGAA 599
Db 2321 ACTCACATGGTGGTGGTAGTTGATGAATCTAGTGGAGATCACGCAACTGCTGCTTGGAA 2380
QY 600 GAACAACCTTAAGGTATTGGGAGATCGATGGGCAACATCTGTAGATGGACAGAACGCCG 659
Db 2381 GAACAACCTTAAGGTATTGGGAGATCGATGGGCAACATCTGTAGATGGACAGAACGCCG 2440
QY 660 TGGGTTCTTTTACAAGAC-----
Db 3521 CACGTCAATGACCTTGTCTGCCAGCTTACCACCTTTGGGCATTCAGCTCTCACCGTATAAC 3580

Db 2441 TGGGTTCTTTTACAAGACCAGCCTGACCTAGCTCCTGGACTGACCACCTATTGGAGCCTCT 2500
QY 678 -----
Db 2501 CCTACTCAGACTGTTACTCTGGTGACACAACCTGTGGTTACTAAGGAAACTGCCATCTCC 2560
QY 678 -----ATCCTTCTCAAATGGCAACGT 698
Db 2561 AAACCTAGAAATGCCATCTTCTTGTATGTTGGAGGTACCTACTCATAGATTACTGCAACAG 2620
QY 699 CTTACTGAAGAACAGTGCCTTTTGTAGTGCATGGCTTTTCAGAAAAAGAAAGATGCAGTGAAC 758
Db 2621 TTCCCTCCTGGACCTGGAAAAAGTTTCTTGCCTGGCTTACAGAAAGCTGAAACAACTGCCAAT 2680
QY 759 AAGATTTCACACAACTGGCTTTA-----AAGATCAAAAATGAAATGTTATCA 803
Db 2681 GTCTTACAGGATGCTACCCGTAAAGGAAAGGCTCCTAGAAAGACTCCAAGGGAGTAAAGAG 2740
QY 804 AGTCITCAAAAACTGGCCGTTTAAAGCGGATCTAGAAAAAGAAAAAGCAATCCATGGGC 863
Db 2741 CTGATGAAACAATGSCAAGACCTCCAAGGTGAAATTTGAAGCTCACACAGATGTTTATCAC 2800
QY 864 AAACCTGTATTCACCTCAAAACAAGATCTTCTTTCAACACTGAAGAAATAAGTCAGTGACCCAG 923
Db 2801 AACCTGGATGAAACACAGCCAAAAAATCCTGAGATCCCTGGAGGTTCGGATGATGCAGTC 2860
QY 924 AAGACGGGAAGCATGGCTGGATAACTTTGCCCGGTGTTGGGATAATTTAGTCCAAAAACTT 983
Db 2861 CTGTTACAAAGACGTTTGGATAACATGAACCTTCAAGTGGAGTGAACCTTCGGAAAAAGTCT 2920
QY 984 GAAAAGAGTACAGCAC-----
Db 2921 CTCAACATTAGGTCCCATTTGGAAGCCAGTTCTTGACCAGTGGAAAGCGTCTGCACCTTTCT 2980
QY 1000 -----
Db 2981 CTGCAGGAACCTTCTGGTGTGGCTACAGCTGAAAGATGATGAATTAAGCCGGCAGGCACCT 3040
QY 1000 -----
Db 3041 ATTGGAGCGCAGCTTCCAGCAGTTCAGAAAGCAGAACGATGTACATAGGGCCTTCAAGAGG 3100
QY 1000 -----
Db 3101 GAATTGAAAACTAAAGAACCTGTAATCATGAGTACTCTTGAGACTGTACGAATATTCTG 3160
QY 1000 -----
Db 3161 ACAGAGCAGCCTTTTGGAAAGGACTAGAGAAACTCTACCAGGAGCCCGCAGAGAGTGCCTCCT 3220
QY 1000 -----
Db 3221 GAGGAGAGAGCCAGAAATGTCACTCGGCTTCTACGAAAGCAGGCTGAGGAGGTCAATACT 3280
QY 1000 -----AGACCCCTT 1007
Db 3281 GAGTGGGAAAAATTTGAACCTGCACCTCCGCTGACTGGCAGAGAAAAATAGATGAGACCCCTT 3340
QY 1008 GAAAGACTCCAGGAACCTTCAAGAGGSCACGGATGAGCTGGACCTCAAGCTCAAGCTGCGCAAGCT 1067
Db 3341 GAAAGACTCCAGGAACCTTCAAGAGGSCACGGATGAGCTGGACCTCAAGCTGCGCAAGCT 3400
QY 1068 GAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGGATCTCCTCATTTGACTCTCTCCAAGAT 1127
Db 3401 GAGGTGATCAAGGGATCCTGGCAGCCCGTGGCGGATCTCCTCATTTGACTCTCTCCAAGAT 3460
QY 1128 CACCTCGAGAAAGTCAAGGCACCTTCAGGAGAAATTTGCCCTCTGAAAGAGAACGTGAGC 1187
Db 3461 CACCTCGAGAAAGTCAAGGCACCTTCAGGAGAAATTTGCCCTCTGAAAGAGAACGTGAGC 3520
QY 1188 CACGTCAATGACCTTGTCTGCCAGCTTACCACCTTTGGGCATTCAGCTCTCACCGTATAAC 1247
Db 3521 CACGTCAATGACCTTGTCTGCCAGCTTACCACCTTTGGGCATTCAGCTCTCACCGTATAAC 3580

QY 1248 CTCAGCACTCTGGAAGACCTGAACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGAC 1307
Db |||||
QY 1308 CGAGTCAGGCAGCTGCATGAAGCCCCACAGGGACTTTGGTCCAGCATCTCAGCACATTCTT 1367
Db |||||
QY 1368 TCCACGTCTGTCCAGGGTCCCTGGGAGAGAGCCATCTCGCCAAACAAAAGTGCCTACTAT 1427
Db |||||
QY 1428 ATCAACCACGAGACTCAAAACAACCTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAG 1487
Db |||||
QY 1488 TCCTTAGCTGACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGA 1547
Db |||||
QY 1548 AGACTGCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTG 1607
Db |||||
QY 1608 GACCAGCACAACTCAAGCAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGT 1667
Db |||||
QY 1668 TTGACCACTATTTATGACCGCCTGGAGCAAGAGCACAAATTTGGTCAACGTCCTCTC 1727
Db |||||
QY 1728 TCGTGGATATGTGTCTGAACCTGGCTGCTGAATGTTATGATACGGGACGAAACAGGAGG 1787
Db |||||
QY 1788 ATCCGTGTCTGTCTTTAAACTGGCATCATTTCCCTGTGTAAGCACACATTTGGAAGAC 1847
Db |||||
QY 1848 AAGTACAGATACCTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGAGCCAGCCGAGG 1907
Db |||||
QY 1908 CTGGGCCCTCCTTCTGCATGATCTTATCCAAATTCCAAGACAGTTGGGTGAAGTTGCATCC 1967
Db |||||
QY 1968 TTTGGGGGCAGTAACATTGAGCCCAAGTGTCGGGA 2001
Db |||||
QY 4301 TTTGGGGGCAGTAACATTGAGCCCAAGTGTCGGGA 4334

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

DM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 07:58:12 ; Search time 1887.39 seconds
(without alignments)
17170.339 Million cell updates/sec

Title: US-09-845-416-14_COPY_1000_3000
Perfect score: 2001
Sequence: 1 ggcagttcattgatggagag.....cattgagccaagtgtccgga 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

total number of hits satisfying chosen parameters: 32308016

Minimum DB seq length: 0
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

- 1: em_estba:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estmu:*
- 5: em_estov:*
- 6: em_estpl:*
- 7: em_estro:*
- 8: em_htc:*
- 9: gb_est1:*
- 10: gb_est2:*
- 11: gb_htc:*
- 12: gb_est3:*
- 13: gb_est4:*
- 14: gb_est5:*
- 15: em_estfun:*
- 16: em_estom:*
- 17: gb_gss:*
- 18: em_gss_hum:*
- 19: em_gss_inv:*
- 20: em_gss_pln:*
- 21: em_gss_vrt:*
- 22: em_gss_fun:*
- 23: em_gss_mam:*
- 24: em_gss_mus:*
- 25: em_gss_other:*
- 26: em_gss_pro:*
- 27: em_gss_rod:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Query %	Length	ID	Description
1	569	28.4	757	9	AL556247	AL556247 AL556247
2	558	27.9	620	14	BQ640063	BQ640063 he23g04.y
3	532.2	26.6	770	12	BG719710	BG719710 602690430
4	511.4	25.6	2135	11	AK013510	AK013510 Mus muscu
5	494.6	24.7	652	10	BB629984	BB629984 BB629984
6	491.6	24.6	704	10	BB610411	BB610411 BB610411

7	467.2	23.3	854	9	AI196693	AI196693 ui53el0.y
8	422.8	21.1	898	12	BF182065	BF182065 601804604
9	410.4	20.5	599	10	BB666688	BB666688 BB666688
10	402.6	20.1	1490	11	BC009242	BC009242 Homo sapi
11	385.2	19.3	684	9	AL641565	AL641565 AL641565
12	381.8	19.1	423	9	AA460476	AA460476 2x61el0.r
13	380.4	19.0	633	9	AL796733	AL796733 AL796733
c	337	16.8	665	10	BE201973	BE201973 fl03h06.x
15	332.6	16.6	2334	11	BC011062	BC011062 Mus muscu
16	330	16.5	608	9	AL672616	AL672616 AL672616
17	329	16.4	532	10	BE334408	BE334408 ug97g06.y
18	325.6	16.3	532	13	BJ036615	BJ036615 BJ036615
19	315.4	15.8	541	13	BI546771	BI546771 603191346
20	310	15.5	488	13	BM488542	BM488542 pgm2n.pk0
21	297.8	14.9	597	9	AL795652	AL795652 AL795652
22	281.8	14.1	573	13	BI446561	BI446561 de24a10.y
23	278.2	13.9	696	13	BJ075057	BJ075057 BJ075057
24	276.6	13.8	647	9	AL652775	AL652775 AL652775
25	256.4	12.8	733	12	BF144271	BF144271 601786885
26	255.4	12.8	772	13	BI250598	BI250598 602993659
27	250	12.5	502	9	AL602076	AL602076 DKFzp313B
c	248.8	12.4	417	12	BF387048	BF387048 UI-R-CA1-
29	248.6	12.4	726	10	BB627285	BB627285 BB627285
30	247.6	12.4	501	9	AA771140	AA771140 vt18b05.r
31	226.4	11.3	784	12	BG212445	BG212445 RST32032
32	225.6	11.3	475	12	BG207912	BG207912 RST27400
33	206.4	10.3	479	14	BQ304046	BQ304046 QV2-BT063
34	190.8	9.5	230	14	N87737	N87737 LL3635F Hum
35	184.8	9.2	307	12	BF226333	BF226333 uz48g06.y
36	173.8	8.7	810	12	BE911128	BE911128 601662540
37	171.2	8.6	835	13	BI553820	BI553820 603190772
c	166.6	8.3	646	13	BI289102	BI289102 UI-R-DK0-
39	163	8.1	714	14	BM931093	BM931093 UI-E-EJ0-
40	162.4	8.1	1002	12	BE798394	BE798394 601583022
41	157.2	7.9	427	12	BE817979	BE817979 CM2-BN027
c	150.4	7.5	515	17	AZ780914	AZ780914 2M0018010
43	144.8	7.2	434	14	BQ375536	BQ375536 PM1-TN012
44	134.6	6.7	770	9	AA392372	AA392372 LD11292.5
45	133.2	6.7	170	9	AA095279	AA095279 13635.seq

ALIGNMENTS

RESULT 1	AL556247	AL556247	757 bp	mrna	linear	EST 16-FEB-2001
LOCUS	AL556247	AL556247	LT1_NFL006_PL2	Homo sapiens	CDNA clone	CS0DK001YB17 5
DEFINITION	AL556247	prime, mRNA sequence.				
ACCESSION	AL556247	AL556247.1	GI:12898746			
VERSION	EST.	human.				
KEYWORDS	human.					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.					
AUTHORS	Li, W.B., Gruber, C., Jessee, J. and Polayes, D.					
TITLE	Full-length cDNA libraries and normalization					
JOURNAL	Unpublished (2001)					
COMMENT	Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - France Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.					
FEATURES	Location/Qualifiers					
source	1..757					
	/organism="Homo sapiens"					
	/db_xref="taxon:9606"					
	/clone="CS0DK001YB17"					
	/clone_lib="LT1_NFL006_PL2"					
	/tissue_type="placenta"					
	/note="Vector: pCMVSPORT 6; Site_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end					

enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : <http://fulllength.invitrogen.com> "

BASE COUNT 190 a 197 c 188 g 182 t

ORIGIN

Query Match	28.4%;	Score 569;	DB 9;	Length 757;
Best Local Similarity	100.0%;	Pred. No. 3.8e-129;		
Matches 569;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;

Qy	1433	CCACGAGACTCAAAACAAC	TTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTT	1492
Db	128	CCACGAGACTCAAAACAAC	TTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTT	187

QY 1493 AGCTGACCTGAATAATGTGAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT 1552
 |||||
 Db 188 AGCTGACCTGAATAATGTGAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT 247

Qy	1553	GCAGAGGCCCTTTGCTTGGATCTCTTGACCTGTGCATGTGATGCCTTGGACCA	1612
Db	248	GCAGAGGCCCTTTGCTTGGATCTCTTGACCTGTGCATGTGATGCCTTGGACCA	307

QY 1613 GCACAACCTCAAGCAAAATGACCAGCCCATGGATATCTCGAGATTATTAATTGTTGAC 1672
|||||
Db 308 GCACAACCTCAAGCAAAATGACCAGCCCATGGATATCTCGAGATTATTAATTGTTGAC 367

Qy	1673	CAC	TAT	TAT	GAC	CGC	CTG	GAG	CAAG	AACA	AAATTGGT	CAACGTCC	CTCTCTGC	GT 1732
Dh	368	CAC	TAT	TAT	GAC	CGC	CTG	GAG	CAAG	AACA	AAATTGGT	CAACGTCC	CTCTCTGC	GT 427

QY 1733 GGATATGTCTGAACTGGCTGCTGAATGTTTATGATACGGGACGAACAGGGAGGATCCG 1792

QY 1793 TGTCTGTCTTTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAGTA 1852

1853 CAGATACCTTTTCAAGCAAGTGGCAAGTTCAACAGGATTTTGTGACCAGCGCAGGCTGGG 1912

Qy 1913 CCTCCTTCTGCATGATCTCTATCCAAATTCCAAGACAGTTGGGTGAAGTTGCATCCTTTGG 1972

008 CCCTCCCTCCGCAIGATTCATCCAAATTCCTCAAGACAGTTGGGTGAAGTTGCATCTTTGG 667
 QY 1973 GGGCAGTAACATTGAGCCAAGTGTCGGA 2001
 |||||

DD 668 GGGCAGTAACATTGAGCCAAAGTGTCGGGA 696
 RESULT 2

BQ640063	LOCUS	BQ640063	620 bp	linear	EST 15-JUL-2002
	DEFINITION	he23g04.y1 Human Retina cDNA (Un-normalized, unamplified): hd/he			

ACCESSION BQ640063
VERSION BQ640063.1 GI:21764522
KEYWORDS EST.
P

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

AUTHORS Wistow, G., Bernstein, S.L., Wyatt, M.K., Ray, S., Behal, A., Touchman, J.W., Bouffard, G., Smith, D. and Peterson, K.

TITLE Expressed sequence tag analysis of human retina for the NEIBank

Project: newbun, an abundant, novel retinal cDNA and alternative

JOURNAL	splicing of other retina-preferred gene transcripts
COMMENT	MOL. VIS. 8 (4), (2002) In press
	Contact: Wistow G

Section on Molecular Structure and Function
National Eye Institute
6/331, NIH, Bethesda, MD 20892-2740, USA
Tel: 301 402 3452

Fax: 301 496 0078
Email: graeme@helix.nih.gov
Plate: 23 row: g column: 04
Seq primer: M13RP1 reverse primer (ABI).

```

FEATURES
  source
    Location/Qualifiers
      1. .620
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="hc23g04"
        /clone_lib="Human Retina cDNA (Un-normalized, unamplified
          ): hd/he"
        /tissue_type="Retina"
        /dev_stage="Adult"
        /lab_host="EMDH10B"

```

/note="Organ: Eye; Vector: pSPORT1; Neural retina tissue was dissected from two 80 year old donors with no observed eye disease. 100ug of total RNA was used for library construction. A directionally cloned cDNA library in the pSPORT1 vector (Life Technologies) was constructed at Bioserve Biotechnology (Laurel MD) essentially following the protocols of the SuperScript Plasmid System full details of which are contained in the manufacturer's Instruction manual (<http://www.lifetech.com/>). First strand synthesis was carried out using a Not I primer-adaptor [5'-pGACTAGTCTAGATCGGAGCGGCGCCC(T)15-3']. EST analysis was performed on the unamplified library at the NIH Intramural Sequencing Center (NISC)."

BASE COUNT	165 a	148 c	144 g	163 t
ORIGIN				

Query Match 27.9%; Score 558; DB 14; Length 620;
Best Local Similarity 100.0%; Pred. No. 1.8e-126;
Matches 558; Conservative 0; Mismatches 0; Indels

QY 1444 AAACAACCTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGA 1503
|||||
Db 1 AAACAACCTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTTAGCTGACCTGA 60

QY 1504 ATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAGGCC 1563

Dh 61 ATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACTGCAGAAGGCC 120

QY	1564	TTTGTCTGGATCTCTTGAGCCCTGTCAGCTGCATGTGATGCCCTTGACCAAGCCTCA	1623
Db	121	TTTTTCCCCTCCCATCTCTGTCAGCCCTGTCAGCTGCATGTGATGCCCTTGACCAAGCCTCA	180

QY 1624 AGCAAAATGACCAGCCCATGGATATCCTGCAGATTATTAATTGTTGACCACATATTATG 1683

QY 1684 ACCGCCCTGGAGCAAGACACACAATTTGGTCAACGTCCCTCTCTGCCGTGGATATGTGTC 1743

1744 TGAAC TGGCTGCTGAATGTTTATGATACGGGACGACAGGAGGATCCCGTGTCCTGTCTT 1803

Qy 1804 TTAAGACTGGCATCATTTCCCTGTGTAAGCACATTTGGAGACAAAGTACAGATACCTTT 1863

301 TTTAAATCTGGCATCATTTCCCTGGTAAAGCACATTTGGAGACAAAGTACAGATACCTTT 420
 QY 1864 TCAAGCAAGTGGCAAGTTCAACAGGATTTGTGACCAGGCAGGCTGGGCTCTTCTGC 1923

421 TCAAGCAAGTGGCAAGTTCACAGGATTTGTGACCAGCGCAGGCTGGGGCCCTCCCTCTGC 480

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS 1 (bases 1 to 1490)
TITLE Strausberg, R.
JOURNAL Direct Submission
Submitted (06-JUN-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Institute for Systems Biology
<http://www.systemsbio.org>
contact: amadnan@systemsbiology.org
Anup Madan, Rachel Dickhoff, Jessica Fahey, Stephanie Ford, Julia Greene, Mark Kettman and Anuradha Madan

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAL Plate: 10 Row: j Column: 10
This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 5032280
This clone has the following problem: incomplete processing.

FEATURES

Location/Qualifiers
1..1490
/organism="Homo sapiens"
/db_xref="LocusID:1756"
/db_xref="taxon:9606"
/clone="IMAGE:3029414"
/tissue_type="Muscle, rhabdomyosarcoma"
/clone_lib="NIH_MGC_17"
/lab_host="DH10B-R"
/note="Vector: pOTB7"

BASE COUNT 505 a 299 c 328 g 358 t
ORIGIN

Query Match 20.1%; Score 402.6; DB 11; Length 1490;
Best Local Similarity 98.8%; Pred. No. 4.1e-88;
Matches 416; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 1 GGCAGTTCATTGATGGAGAGTGAAGTAAACCTGGACCGTATTCAAACAGCTTTAGAAGAA 60
Db 1069 GGCAGTTCATTGATGGAGAGTGAAGTAAACCTGGACCGTATTCAAACAGCTTTAGAAGAA 1128
QY 61 GTATTATCGTGGCTTCTTCTGCTGAGGACACATTCGAAGCACAAAGGAGAGATTCTAAT 120
Db 1129 GTATTATCGTGGCTTCTTCTGCTGAGGACACATTCGAAGCACAAAGGAGAGATTCTAAT 1188
QY 121 GATGTGGAAGTGGTGAAGACCAGTTTCATACATGAGGGGTACATGATGGATTGACA 180
Db 1189 GATGTGGAAGTGGTGAAGACCAGTTTCATACATGAGGGGTACATGATGGATTGACA 1248
QY 181 GCCCATCAGGGCCGGGTTGGTAATATTCTACAATTGGGAAGTAAGCTGATTGGAACAGGA 240
Db 1249 GCCCATCAGGGCCGGGTTGGTAATATTCTACAATTGGGAAGTAAGCTGATTGGAACAGGA 1308
QY 241 AAATTATCAGAAGATGAAGAAACTGAAGTACAAGACGACAGATGAATCTCTAAATTCAAGA 300
Db 1309 AAATTATCAGAAGATGAAGAAACTGAAGTACAAGACGACAGATGAATCTCTAAATTCAAGA 1368
QY 301 TGGGAATGCCTCAGGGTAGCTAGCATGGAAACAAAGCAATTTACATAGAGTTTAAATG 360
Db 1369 TGGGAATGCCTCAGGGTAGCTAGCATGGAAACAAAGCAATTTACATAGAGTTTAAATG 1428
QY 361 GATCTCCAGAATC-GAAACTGAAAGAGTTGAATGACTGGCTAACCAAAAACAGAGAAGA 419
Db 1429 GATCTCCAGAATCAGAAACTGAAAGAGTTGAATGACTGGCTAACCAAAAACAGAGAAGA 1488
QY 420 A 420

Db 1489 A 1489

RESULT 11
AL641565

LOCUS AL641565 XGC-neurula silurana tropicalis cDNA clone TNeu012d20 5',
DEFINITION mRNA sequence.
ACCESSION AL641565
VERSION AL641565.1 GI:16793690
KEYWORDS EST.
SOURCE western clawed frog.
ORGANISM Silurana tropicalis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Amphibia; Batrachia; Anura; Mesobatrachia; Pipidea; Pipidae; Xenopodinae; Silurana.
1 (bases 1 to 684)
Huckle, E., Taylor, R., Ashurst, J.L., Zorn, A.M. and Rogers, J.
Sanger Xenopus tropicalis EST project 2001 (10_2001)
Unpublished (2001)
Contact: Huckle E
Sanger Centre

Hinxton, Cambridgeshire, CB10 1SA, UK
Email: trop@sanger.ac.uk
Sanger Xenopus tropicalis EST project 2001
TROPICALIS_SEQUENCE_ID: TNeu012d20.sp6
Sequencing primer: SP6

This sequence is from a Xenopus Gene Collection (XGC) library
constructed by Aaron M. Zorn.

Location/Qualifiers
1..684
/organism="Silurana tropicalis"
/db_xref="taxon:8364"
/clone="TNeu012d20"
/clone_lib="XGC-neurula"
/dev_stage="neurula"
/lab_host="Escherichia coli DH10B"
/note="Vector: pCS107; Site_1: EcoRI; Site_2: NotI; cDNA was oligo dT primed from 5ug of poly A+ RNA from neurula. EcoRI-NotI cut cDNA was then ligated into pCS107 with EcoRI at the 5' end and NotI at the 3' end."

BASE COUNT 168 a 175 c 172 g 169 t
ORIGIN

Query Match 19.3%; Score 385.2; DB 9; Length 684;
Best Local Similarity 84.7%; Pred. No. 6.3e-84;
Matches 432; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 1433 CCACGAGACTCAAACAACCTTGCTGGGACCATCCCAAAATGACAGAGCTCTACCAGTCTTT 1492
Db 175 CCATGAGACACAAACCACTGCTGGGATCATCCCAAAATGACAGAATTATACCAATCTTT 234
QY 1493 AGCTGACCTGAATAATGTCAGATTCTCAGCTTATAGGACTGCCATGAAACTCCGAAGACT 1552
Db 235 AGCTGACCTGAACAATGTGCGATTCTCAGCATAACAGAACTGCCATGAAGCTAAGGAGATT 294
QY 1553 GCAGAAGGCCCTTTGCTTGGATCTCTTGAGCCTGTCAGCTGCATGTGATGCCCTTGGACCA 1612
Db 295 GCAAAAGGCCCTTGCTTGGATTGCTAGGGCTGTCTGCAGCTTGTGAAGCCTTGGACCA 354
QY 1613 GCACAACCTCAAGCAAAATGACCAGCCCATGGATTCCTGCAGATTATTAAATGTTTGAC 1672
Db 355 GCACAACCTGAAGCAGAAATGACCAGCTGTATGGACATCTGCAGATTATTAAATGTTTGAC 414
QY 1673 CACTATTTATGACCGCTGGAGCAAGAGCACACAATTTGGTCAACGTCCCTCTCTCGCT 1732
Db 415 CACAATTTATGATCGACTGGAGCAAGAGCACACAATTTGGTGAACGTTCTCTCTCGCT 474
QY 1733 GGATATGTGTCTGAAGCTGGCTGCTGAATGTTTATGATACGGGACGACAGGAGGATCCG 1792
Db 475 GGACATGTGCTCAACTGGCTGCTGAATGTTTATGACACGGGTCGACGGGACGTATACG 534

AUTHORS Strausberg, R.
TITLE Direct Submission
JOURNAL Submitted (25-JUL-2001) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>
COMMENT Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Gilbert Smith, Ph.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hqsc.bcm.tmc.edu/cdna/>
Contact: amg@bcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Hulyk, S.W., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>
Series: IRAK Plate: 23 Row: j Column: 4
This clone was selected for full length sequencing because it passed the following selection criteria: Hexamer frequency ORF analysis, Similarity but not identity to protein
This clone has the following problem: incomplete processing.

FEATURES

Source
1. .2334
/organism="Mus musculus"
/db_xref="taxon:10090"
/map="FVB/N"
/clone="IMAGE:3979320"
/tissue_type="Mammary tumor. Metallothionien-TGF alpha model. 10 month old virgin mouse. Taken by biopsy."
/clone_lib="NCI_CGAP_Mam1"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
BASE COUNT 698 a 512 c 638 g 486 t
ORIGIN

Query Match 16.6%; Score 332.6; DB 11; Length 2334;
Best Local Similarity 59.0%; Pred. No. 7.8e-71;
Matches 589; Conservative 0; Mismatches 409; Indels 1; Gaps 1;

QY 4 AGTTCATTGATGGAGAGTGAACTAAACCTGGACCGTTATCAAACAGCTTTAGAAGAAGTA 63
Db 1042 AGCACCGTCACTGAAGTGGACATGGATTGGACAGCTACCAGATAGCGCTAGAGGAAGTG 1101
QY 64 TTATCGTGGCTTCTTCTGCTGAGGACACACATTGCAAGCACAAAGGAGAGATTCTTAATGAT 123
Db 1102 CTGACGTGGCTGCTGCCGCGGAGGACACAGTTCCAGGAGCAAGATGACATTCTGATGAT 1161
QY 124 GTGGAAGTGGTAAAGACCACTTTCATACCTCATGAGGGGTACATGATGGATTGACAGCC 183
Db 1162 GTCGAAGAAGTCAAAGAGCAGTTTGCTACCCATGAAACTTTTATGATGGAGCTGACAGCA 1221
QY 184 CATCAGGGCCGGTGGTAATATTTCTACAATTGGGAAGTAAGCTGATTGGAACAGGAAAA 243
Db 1222 CACCAGAGCAGCGTGGGGAGCGTCTGCAGGGCTGGCAACCAGCTGATGACACAAGGGACT 1281
QY 244 TTATCAGAAGATGAAGAACTGAAGTACAAGAGCAGATGAATCTCCTAAATTCGAAGATGG 303
Db 1282 CTGTCAGAGGAGGAGGAGTTTGAGATCCAGGAACAGATGACCTTGCTGAATGCAAGGTGG 1341
QY 304 GAATGCCTCAGGGTAGCTAGATGGAAAAACAAAGCAATTTACATAGAGTTTAAATGGAT 363
Db 1342 GAGGCGCTCCGGTGGAGAGCATGGAGAGGCGAGTCCCGGCTGCACGCGCTCTGATGGAG 1401
QY 364 CTCCAGAATCGA-AACTGAAAGAGATTGAATGACTGGCTAACAAAAACAGAAAGAAACA 422
Db 1402 CTGCAGAAGAAACAGCTGCACAGACTCTCAAGCTGGCTGGCCCTCACAGAAGAGCGCATT 1461

QY 423 AGGAAAATGGAGGAGAGCCCTCTTTGGACCTGATCTTTGAAGACCTTAAACGCCAAGTACAA 482
Db 1462 CAGAAGATGGAGAGCCCTCCCGCTGGGTGATGACCTGCCCTCCCTGCAGAGAGTGTCTCAA 1521
QY 483 CAACATAAGGTGCTTCAAGAAAGATCTAGAACAAAGAACAAAGTCAGGGTCAATTCTCTCACT 542
Db 1522 GAACATAAAAAGTTTGCAAAATGACCTTGAAGCTGAACAGGTGAAGTTAAATTCCTTAACT 1581
QY 543 CACATGGTGGTGTAGTTGATGAATCTAGTGGAGATCACGCAACTGCTGCTTTGGAAGAA 602
Db 1582 CACATGGTGGTGTAGTTGATGAATAAACAGTGGGAGAGTGCACAGCTCTCTGGAAGAT 1641
QY 603 CAACTTAAGGTATTGGGAGATCGATGGGCAACATCTGTAGATGGACAGAAGACCGCTGG 662
Db 1642 CAGTTACAGAAAAGTGGGTGAGCGCTGGACAGCTGTATGCCGCTGGACTGAAGAACGTTGG 1701
QY 663 GTTCTTTTACAAGACATCCTTCTCAAAATGGCAACGCTCTTACTGAAGAACAGTGCCTTTT 722
Db 1702 AACAGGTTGCAAGAAATCAGTATTCTGTGGCAGGAATTATTGGAAGAGCAGTGTCTGTTG 1761
QY 723 AGTGCATGGCTTTTCAGAAAAAAGAGATGCAGTGAACAAGATTTCACACAACTGGCTTTAA 782
Db 1762 GAGGCTTGGCTCACCCGAAAAGGAAGAGGCTTTGAATAAAGTTCAAACCCAGCAACTTTAA 1821
QY 783 GATCAAAATGAAATGTTATCAAGTCTTCAAAAACCTGGCCGCTTTTAAACCGGATCTAGAA 842
Db 1822 GACCAGAAGGAACTAAGTGTCAAGTGTCCGGCGTCTGGCTATATTGAAGGAAGACATGGAA 1881
QY 843 AAGAAAAAGCAATCCATGGGCAAACTGTATTCACTCAAACAAGATCTTCTTTCAACACTG 902
Db 1882 ATGAAGAGGCAGACTCTGGATCAACTGAGTGAGATTGGCCAGGATGTGGGCCAATTACTC 1941
QY 903 AAGAATAAGTCAAGTACCCAGAAAGACGGAAGCATGGCTGGATAACTTTGCCCGGTGTTGG 962
Db 1942 AGTAATCCCAAGGCATCTAAGAAGATGAACAGTGAACAGTGAAGGAGCTTAACACAGAGATGG 2001
QY 963 GATAATTTAGTCCAAAAACTTGAAAAAGAGTACAGCACAG 1001
Db 2002 GATTCTCTGGTTCAGAGACTCGAAGACTCTTCTTAACCAG 2040

Search completed: April 25, 2003, 17:39:40
Job time : 1914.39 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

DM nucleic - nucleic search, using sw model

Run on: April 25, 2003, 08:10:12 ; Search time 45.5741 Seconds
(without alignments)
13465.096 Million cell updates/sec

Title: US-09-845-416-14_COPY_1000_3000
Perfect score: 2001
Sequence: 1 ggcagttcattgatggagag.....cattgagccaagtgtccgga 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 875442

Minimum DB seq length: 0
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/1/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/1/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/1/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/1/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/1/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
C 1	42.4	2.1	2704	4	US-08-857-076-44 Sequence 44, Appl
C 2	42.4	2.1	3499	4	US-08-857-076-43 Sequence 43, Appl
C 3	39.4	2.0	2223	1	US-08-257-073-4 Sequence 4, Appl
C 4	38.8	1.9	289	4	US-09-007-005-17 Sequence 17, Appl
C 5	38.8	1.9	289	4	US-09-244-796-17 Sequence 17, Appl
C 6	38.8	1.9	1821	4	US-08-477-831C-1 Sequence 1, Appl
C 7	38.8	1.9	1885	4	US-08-477-831C-9 Sequence 9, Appl
C 8	38.8	1.9	1896	4	US-08-477-831C-10 Sequence 10, Appl
C 9	38.8	1.9	1961	4	US-08-477-831C-8 Sequence 8, Appl
C 10	38.8	1.9	2968	4	US-08-477-831C-13 Sequence 13, Appl
C 11	38.8	1.9	3044	4	US-08-477-831C-12 Sequence 12, Appl
C 12	38.2	1.9	1848	4	US-09-134-001C-447 Sequence 447, Appl
C 13	37.4	1.9	2082	3	US-08-985-335-4 Sequence 4, Appl
C 14	37.4	1.9	2082	4	US-09-410-372-4 Sequence 4, Appl
C 15	37.2	1.9	3902	4	US-08-961-527-212 Sequence 212, Appl
C 16	36	1.8	608	4	US-09-385-982-236 Sequence 236, Appl
C 17	36	1.8	2763	1	US-08-248-466B-2 Sequence 2, Appl
C 18	35.8	1.8	633	4	US-09-134-001C-578 Sequence 578, Appl
C 19	35.6	1.8	2447	2	US-09-014-969-14 Sequence 14, Appl
C 20	35.6	1.8	4868	1	US-08-139-937-12 Sequence 12, Appl
C 21	35.6	1.8	4868	5	PCT-US93-11310-12 Sequence 12, Appl
C 22	35.2	1.8	1086	1	US-08-415-751-27 Sequence 27, Appl
C 23	35.2	1.8	1086	1	US-08-415-751-28 Sequence 28, Appl
C 24	35.2	1.8	1089	3	US-08-642-807A-27 Sequence 27, Appl
C 25	35.2	1.8	1845	4	US-08-887-534A-22 Sequence 38, Appl
C 26	35.2	1.8	3466	1	US-08-468-036-38 Sequence 38, Appl
C 27	35.2	1.8	3466	2	US-08-376-843-38 Sequence 38, Appl

C 28	34.8	1.7	366	4	US-09-134-001C-997 Sequence 997, Appl
C 29	34.6	1.7	2277	1	US-08-676-967-5 Sequence 5, Appl
C 30	34.6	1.7	2277	1	US-08-676-974-5 Sequence 5, Appl
C 31	34.6	1.7	2277	2	US-09-098-487-5 Sequence 5, Appl
C 32	34.2	1.7	621	4	US-09-328-111-204 Sequence 204, Appl
C 33	34.2	1.7	1620	4	US-08-858-207A-56 Sequence 56, Appl
C 34	34.2	1.7	2800	2	US-08-874-138-1 Sequence 1, Appl
C 35	34.2	1.7	2800	2	US-08-874-138-5 Sequence 5, Appl
C 36	34.2	1.7	2800	4	US-08-879-941-1 Sequence 1, Appl
C 37	34.2	1.7	2800	4	US-08-879-941-3 Sequence 3, Appl
C 38	34.2	1.7	2800	4	US-09-747-116-1 Sequence 1, Appl
C 39	34.2	1.7	2800	4	US-09-747-116-3 Sequence 3, Appl
C 40	34	1.7	248	4	US-09-007-005-32 Sequence 32, Appl
C 41	34	1.7	248	4	US-09-244-796-32 Sequence 32, Appl
C 42	34	1.7	277	4	US-09-007-005-3 Sequence 3, Appl
C 43	34	1.7	277	4	US-09-244-796-3 Sequence 3, Appl
C 44	34	1.7	2235	4	US-09-153-804-2 Sequence 2, Appl
C 45	34	1.7	2873	4	US-08-630-915A-193 Sequence 193, Appl

ALIGNMENTS

RESULT 1
US-08-857-076-44/c
; Sequence 44, Application US/08857076C
; Patent No. 6225120
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Kimura, Koutarou
; APPLICANT: Patterson, Garth
; APPLICANT: Ogg, Scott
; APPLICANT: Paradis, Suzanne
; APPLICANT: Tissenbaum, Heidi
; APPLICANT: Morris, Jason
; APPLICANT: Kowek, Allison
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351001
; CURRENT APPLICATION NUMBER: US/08/857,076C
; CURRENT FILING DATE: 1997-05-15
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 44
; LENGTH: 2704
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-08-857-076-44

Query Match 2.1%; Score 42.4; DB 4; Length 2704;
Best Local Similarity 48.4%; Pred.No. 0.011;
Matches 118; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY	196	GTGGTAATATTCTACAATTGGGAAGTAAGCTGATTGGAACAGGAAAAATATCAGAAGAT	255
Db	2116	GCTAGACATTTGCTACCAGGGGGGAAAAAGAGGAAAAATTAAGAAATTACAAGAGAAAT	2057
QY	256	GAAGAACTGAAGTACAAGAGCAGATGAATCTCTTAATTCAGATGGGAATGCCTCAGG	315
Db	2056	CAAGAAATCAAAATTTTGGAAAAATTCAAATTAAGAAAGACAGATTGGGGAATGTGTTTA	1997
QY	316	GTAGCTAGCATGGAAAAACAAGCAATTTACATAGAGTTTTTAATGGATCTCCAGAATCGA	375
Db	1996	GCAGTTTTGATTGCGAAACAAATTTACAGCGAAAAAGATGGGATAGAGGTAGCATTTAAA	1937
QY	376	AACTGAAAGAGTTGAATGACTGGCTAAACAAAAACAGAAAGAAATGGAGG	435
Db	1936	GAGTGAAGAACCTGGATGACGGGAGAGAGAGAGAGAAAAACAAAAAGAGGAT	1877
QY	436	AAGA 439	
Db	1876	AATA 1873	


```
RESULT 2
US-08-857-076-43/c
; Sequence 43, Application US/08857076C
; Patent No. 6225120
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Kimura, Koutarou
; APPLICANT: Patterson, Garth
; APPLICANT: Ogg, Scott
; APPLICANT: Paradis, Suzanne
; APPLICANT: Tissenbaum, Heidi
; APPLICANT: Morris, Jason
; APPLICANT: Koweek, Allison
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351001
; CURRENT APPLICATION NUMBER: US/08/857,076C
; CURRENT FILING DATE: 1997-05-15
; NUMBER OF SEQ ID NOS: 114
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 3499
; TYPE: DNA
; ORGANISM: Caenorhabditis elegans
US-08-857-076-43

Query Match      2.1%; Score 42.4; DB 4; Length 3499;
Best Local Similarity 48.4%; Pred. No. 0.014;
Matches 118; Conservative 0; Mismatches 126; Indels 0; Gaps 0;

QY 196 GTTGGTAATATTCTACAAATGGGAAGTAAGCTGATTGGAACAGGAAAAAATTATCAGAAGAT 255
Db 2911 GCTAGACATTTGCTACCGGGGAAAAAAGAGGAAAAATTAAAGAAATTACAAGAGAAAT 2852

QY 256 GAAGAACTGAAGTACAAGACAGACAGATGAATCTCCTAAATTCAGATGGGAATGCCTCAGG 315
Db 2851 CAAGAAATCAAAATTTTGAAGAAATTCATTAAGAAAGACAGATTTGGGGAATCTGTTTA 2792

QY 316 GTAGCTAGCATGGAAAAACAAAGCAATTTACATAGAGTTTTTAATGGATCTCCAGAATCGA 375
Db 2791 GCAGTTTGTGTCGGAACAAATTTACAGCGAAAAAGATGGGATAGAGGTAGCATTTAA 2732

QY 376 AACTGAAAGAGTTGATGACTGGCTAACAAAAACAGAAAGAAAGACAAAGGAAAAATGGAGG 435
Db 2731 GAGTGAAGAACCTGGATGACGGGAGAGAGAGAGAGAAAAAACAAGAAAAAGAGGT 2672

QY 436 AAGA 439
Db 2671 AATA 2668

RESULT 3
US-08-257-073-4
; Sequence 4, Application US/08257073
; Patent No. 5766597
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: de Taisne, Charles
; APPLICANT: Tine, John A.
; TITLE OF INVENTION: MALARIA RECOMBINANT POXVIRUS VACCINE
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue, 25th Floor
; CITY: New York
; STATE: New York
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/257,073
FILING DATE: 09-JUN-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/075,783
FILING DATE: 11-JUN-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/852,305
FILING DATE: 18-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/672,183
FILING DATE: 20-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2570
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066 CURTMS
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2223 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-257-073-4

Query Match      2.0%; Score 39.4; DB 1; Length 2223;
Best Local Similarity 51.4%; Pred. No. 0.086;
Matches 91; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 368 AGAATCGAACTGAAAGAGTTGAATGACTGGCTAACAAAAACAGAAAGAAACAAAGGAA 427
Db 2034 AAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAA 2093

QY 428 AATGGAGGAAGAGCCTCTTTGGACCTGTATCTTGAAGACCTTAAACGCCCAAGTACAACA 487
Db 2094 AAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAA 2153

QY 488 TAAGGTGCTTCAAGAAGATCTAGAACAAGAACAAAGTCAAGGTCAATTTCTCACTCA 544
Db 2154 AGAAGAAGAACACAGAAAGAAAGAAAGAAAGAAATAGTACCAGAAAAATTTGACAACTGA 2210

RESULT 4
US-09-007-005-17
; Sequence 17, Application US/09007005B
; Patent No. 6258558
; GENERAL INFORMATION:
; APPLICANT: Szostak, Jack W.
; APPLICANT: Roberts, Richard W.
; APPLICANT: Liu, Rihe
; TITLE OF INVENTION: SELECTION OF PROTEINS USING RNA-PROTEIN
; FILE REFERENCE: 00786/350003
; CURRENT APPLICATION NUMBER: US/09/007,005B
; CURRENT FILING DATE: 1998-01-14
; EARLIER APPLICATION NUMBER: 60/035,963
; EARLIER FILING DATE: 1997-01-27
; EARLIER APPLICATION NUMBER: 60/064,491
; EARLIER FILING DATE: 1997-11-06
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 289
; TYPE: RNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Translation template
; FEATURE:
```


409 CAGAGAAAGAACAAAGGAAAATGGAGGAAGAGCCCTCTTGGACCTGATCTCTGAAGACCTAA 46

QY 390 AATGACTGGCTAACACAAACAGAGAAAGAAACAAAGGAAATGGAGGAGAGCCCTCTTGGA 449
| | | | |
Db 1513 AATGAACATGGTGCAACTCAGGAGCAGCTAAATAAAATCAGAGACTCCTATGCACAGCTA 1572
| | | | |
QY 450 CCTGATCTTGAAGACCTTAAACGCCCAAGTACAACAACATAAGGTGCTTCAAGAAGA 505
| | | | |
Db 1573 CTGGTCAACCAAGACCTTAAAGCAAAAATCAAAACATGTTGTGAAATGAAAGATGA 1628
| | | | |

RESULT 7
US-08-477-831C-9
; Sequence 9, Application US/08477831C
; Patent No. 6429291
; GENERAL INFORMATION:
; APPLICANT: TURLEY, EVA A.
; APPLICANT: SHUWEN, ZHANG
; APPLICANT: ENTWISTLE, JOYCELYN
; TITLE OF INVENTION: HYALURONAN RECEPTOR PROTEIN
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 AVENUE OF THE AMERICAS
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10020-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Rel. #1.0, ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,831C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: PIERRI, MARGARET A.
; REGISTRATION NUMBER: 30,709
; REFERENCE/DOCKET NUMBER: SIM-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1885 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; DESCRIPTION: /desc = "RHAMM IB"

US-08-477-831C-9
Query Match 1.9%; Score 38.8; DB 4; Length 1885;
Best Local Similarity 48.0%; Pred. No. 0.12;
Matches 142; Conservative 0; Mismatches 152; Indels 2; Gaps 1;

QY 210 ACAATTGGGAAGTAAGCTGATTGGAACAGGAAATATCAGAAGATGAAGAACTGAAGT 269
| | | | |
Db 1399 AGACTTTAGGAAGCAGCTGGAAGAGAGAAAGGAAAGAAACAGACAGAGAGAAATGTAAT 1458
| | | | |
QY 270 ACAAGACAGATGAATCTCCTAAATTCAGATGGGAATGCCCTCAGGGTAGCTAGCATGGA 329
| | | | |
Db 1459 GACAGAAATTAACCATGGAATTAAT--AAATGGCGTCTCCTATATGATGAATATATGAA 1516
| | | | |
QY 330 AAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCGAAACTGAAAGAGTTG 389
| | | | |
Db 1517 AAAACTAAACCTTTTCAGCAACAACACTGGATGGCTTTGAAGCCGAGAAACAGGCATTGTTG 1576
| | | | |
QY 390 AATGACTGGCTAACAAAAACAGAGAAAGAAAGCAAGGAAATGGAGGAGAGAGCCCTCTTGGA 449
| | | | |
Db 1577 AATGAACATGGTGCAACTCAGGAGCAGCTAAATAAAATCAGAGACTCCTATGCACAGCTA 1636
| | | | |

QY 450 CCTGATCTTGAAGACCTTAAACGCCCAAGTACAACAACATAAGGTGCTTCAAGAAGA 505
| | | | |
Db 1637 CTGGTCAACCAAGACCTTAAAGCAAAAATCAAAACATGTTGTGAAATGAAAGATGA 1692
| | | | |

RESULT 8
US-08-477-831C-10
; Sequence 10, Application US/08477831C
; Patent No. 6429291
; GENERAL INFORMATION:
; APPLICANT: TURLEY, EVA A.
; APPLICANT: SHUWEN, ZHANG
; APPLICANT: ENTWISTLE, JOYCELYN
; TITLE OF INVENTION: HYALURONAN RECEPTOR PROTEIN
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 AVENUE OF THE AMERICAS
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10020-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Rel. #1.0, ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/477,831C
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: PIERRI, MARGARET A.
; REGISTRATION NUMBER: 30,709
; REFERENCE/DOCKET NUMBER: SIM-10
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 10:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1896 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; DESCRIPTION: /desc = "RHAMM I-2a"

US-08-477-831C-10
Query Match 1.9%; Score 38.8; DB 4; Length 1896;
Best Local Similarity 48.0%; Pred. No. 0.12;
Matches 142; Conservative 0; Mismatches 152; Indels 2; Gaps 1;

QY 210 ACAATTGGGAAGTAAGCTGATTGGAACAGGAAATATCAGAAGATGAAGAACTGAAGT 269
| | | | |
Db 1410 AGACTTTAGGAAGCAGCTGGAAGAGAGAAAGGAAAGAAACAGACAGAGAGAAATGTAAT 1469
| | | | |
QY 270 ACAAGACAGATGAATCTCCTAAATTCAGATGGGAATGCCCTCAGGGTAGCTAGCATGGA 329
| | | | |
Db 1470 GACAGAAATTAACCATGGAATTAAT--AAATGGCGTCTCCTATATGATGAATATATGAA 1527
| | | | |
QY 330 AAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCGAAACTGAAAGAGTTG 389
| | | | |
Db 1528 AAAACTAAACCTTTTCAGCAACAACACTGGATGGCTTTGAAGCCGAGAAACAGGCATTGTTG 1587
| | | | |
QY 390 AATGACTGGCTAACAAAAACAGAGAAAGAAAGCAAGGAAATGGAGGAGAGAGCCCTCTTGGA 449
| | | | |
Db 1588 AATGAACATGGTGCAACTCAGGAGCAGCTAAATAAAATCAGAGACTCCTATGCACAGCTA 1647
| | | | |
QY 450 CCTGATCTTGAAGACCTTAAACGCCCAAGTACAACAACATAAGGTGCTTCAAGAAGA 505
| | | | |
Db 1648 CTGGTCAACCAAGACCTTAAAGCAAAAATCAAAACATGTTGTGAAATGAAAGATGA 1703
| | | | |

APPLICANT: TURLEY, EVA A.
APPLICANT: SHUWEN, ZHANG
APPLICANT: ENTWISTLE, JOYCELYN
TITLE OF INVENTION: HYALURONAN RECEPTOR PROTEIN
NUMBER OF SEQUENCES: 60
CORRESPONDENCE ADDRESS:
ADDRESSEE: FISH & NEAVE
STREET: 1251 AVENUE OF THE AMERICAS
CITY: NEW YORK
STATE: NEW YORK
COUNTRY: U.S.A.
ZIP: 10020-1104

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/477,831C
FILING DATE: 07-JUN-1995
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: PIERRI, MARGARET A.
REGISTRATION NUMBER: 30,709
REFERENCE/DOCKET NUMBER: SIM-10
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000

TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 2968 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

```

;      TOPOLOGY: linear
;      MOLECULE TYPE: CDNA
;      DESCRIPTION: /desc = "RHAMM IB cDNA (includes 3' sequence)"
US-08-477-831C-13

```

```
Query Match          1.9%; Score 38.8; DB 4; Length 2968;
Best Local Similarity 48.0%; Pred. No. 0.16;
Matches 142; Conservative 0; Mismatches 152; Indels 2
```

QY	210	ACAAATTGGGAAGTAAGCTGATTTGGAAACAGGAAAAATATCAGAAGATGAAGAAAACTGAAGT	269
Db	1399	AGACTTTAGGAAGCAGCTGGAAGAGAAAGGAAAAAAGAACAGACAGAGAAAGAAAAATGTAAAT	1458
QY	270	ACAAGAGCAGATGAATCTCCTAAATTCAGAGATGGGAATGCCCTCAGGGTAGCTAGCATGGA	329
Db	1459	GACAGAAATTAACCATGGAAATTAAT--AAATGGCGTCTCCTATATGATGAACATATATGAA	1516
QY	330	AAAAACAAAGCAATTTACATAGAGTTTTTAATGGATCTCCAGAATCGAAACTGAAAAGATTG	389
Db	1517	AAAACTAAACCTTTTCAGCAACAACCTGGATGCCCTTTGAAGCCGAGAAACAGGCATTGTTG	1576
QY	390	AATGACTGGCTAAACAAAAACAGAAGAAAGAACAGGAAAAATCGAGGAAGAGCCCTCTTGGA	449
Db	1577	AATGAACATGGTGCAACTCAGGAGCAGCTAAATAAAATCAGAGACTCCTATGCACAGCTA	1636
QY	450	CCTGATCTTGAAGACCTTAAACGCCCAAGTACAACAACATAAAGTGCTTCAAGAAGA	505
Db	1637	CTTCCCTGACCAAGACCTTAAAGCAAAAAATCAAAACATGTTGTGAAATTTGAAAGATGA	1692

```

RESULT 11
US-08-477-831C-12
; Sequence 12, Application US/08477831C
; Patent No. 6429291
; GENERAL INFORMATION:
; APPLICANT: TURLEY, EVA A.
; APPLICANT: SHUWEN, ZHANG
; APPLICANT: ENTWISTLE, JOYCELYN
; TITLE OF INVENTION: HYALURONAN RECEPTOR PROTEIN
; NUMBER OF SEQUENCES: 60

```

```

;
; APPLICANT: TURLEY, EVA A.
;
; APPLICANT: SHUWEN, ZHANG
;
; APPLICANT: ENTWISTLE, JOYCELYN
;
; TITLE OF INVENTION: HYALURONAN RECEPTOR PROTEIN
;
; NUMBER OF SEQUENCES: 60
;

```


;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: FISH & NEAVE
;; STREET: 1251 AVENUE OF THE AMERICAS
;; CITY: NEW YORK
;; STATE: NEW YORK
;; COUNTRY: U.S.A.
;; ZIP: 10020-1104
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Rel. #1.0, ASCII
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/477,831C
;; FILING DATE: 07-JUN-1995
;; CLASSIFICATION: 514
;; ATTORNEY/AGENT INFORMATION:
;; NAME: PIERRI, MARGARET A.
;; REGISTRATION NUMBER: 30,709
;; REFERENCE/DOCKET NUMBER: SIM-10
;; TELEPHONE: 212-596-9000
;; TELEFAX: 212-596-9090
;; INFORMATION FOR SEQ ID NO: 12:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 3044 base pairs
;; TYPE: nucleic acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: cDNA
;; DESCRIPTION: /desc = "RHAMM IA cDNA (includes 3' sequence)"
;; IS-08-477-831C-12

Query Match 1.9%; Score 38.8; DB 4; Length 3044;
Best Local Similarity 48.0%; Pred. No. 0.17;
Matches 142; Conservative 0; Mismatches 152; Indels 2; Gaps 1;
Y 210 ACAATTGGGAAGTAAGCTGATTGGAACAGGAAATTTATCAGAGATGAAGAACTGAAGT 269
b 1475 AGACTTTAGGAAGCAGCTGGAGAGAGAAAGAAAGAACAGCAGAGAGAAAGAAATGTAAT 1534
Y 270 ACAAGAGCAGATGAATCTCTAAATTTCAAGATGGGAATGCCTCAGGGTAGCTAGCATGGA 329
b 1535 GACAGAATTAACCATGGAAATTAAT--AAATGGCGTCTCCTATATGATGAACATATGAA 1592
Y 330 AAAACAAAGCAATTTACATAGAGTTTAAATGGATCTCCAGAAATCGAAACTGAAAGAGTTG 389
b 1593 AAAACTAAACCTTTTCAGCAACAACTGGATGCCTTTGAAGCCGAGAAACAGGCATTGTTG 1652
Y 390 AATGACTGGCTAACAAAACAGAGAAAGAAAGAAAGAAATGGAGGAGAGCCCTCTTGA 449
b 1653 AATGAACATGGTGCAACTCAGGAGCAGCTAAATAAATCAGAGACTCCTATGCACAGCTA 1712
Y 450 CCTGATCTTGAAGACCTTAAACGCCCAAGTACAAACATAGAGTGGCTTCAAGAAGA 505
b 1713 CTTGGTCACCAAGACCTTAAAGCAAAAATCAACATGTTGTGAAATTTGAAAGATGA 1768

RESULT 12
S-09-134-001C-447
Sequence 447, Application US/09134001C
Patent No. 6380370
GENERAL INFORMATION:
APPLICANT: Lynn Doucette-Stamm et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS
TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC-007
CURRENT APPLICATION NUMBER: US/09/134,001C
CURRENT FILING DATE: 1998-08-13
PRIOR APPLICATION NUMBER: US 60/064,964
PRIOR FILING DATE: 1997-11-08
PRIOR APPLICATION NUMBER: US 60/055,779
PRIOR FILING DATE: 1997-08-14

;; NUMBER OF SEQ ID NOS: 5674
;; SEQ ID NO 447
;; LENGTH: 1848
;; TYPE: DNA
;; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-447
Query Match 1.9%; Score 38.2; DB 4; Length 1848;
Best Local Similarity 48.0%; Pred. No. 0.18;
Matches 109; Conservative 0; Mismatches 118; Indels 0; Gaps 0;
QY 752 AGTGAACAAGATTACACAACTGGCTTTAAAGATCAAAAATGAAATGTTATCAAGTCTTCA 811
Db 207 AGTTGAAATTTTCTCAATAATCACCTTTTATCGTTAAAGTAGCGGAAAAAGTGCCTT 266
QY 812 AAAAAGTGGCGCTTTTAAAGCGGATCTAGAAAAAGAAAGCAATCCATGGGCAAACTGTA 871
Db 267 AAAAAGATACGTTTTTACTGGTGACATAGTATTAATAAATAAATAAATTTCTCAATCAAT 326
QY 872 TTCACCTCAAAACAAGATCTTCTTCAACACTGAAGAATAAGTCAGTGACCCAGAGACCGGA 931
Db 327 TGAAAGTTTAGAAGAACTTCTTCAATACTAAGGAAATATTTTCATATCAACATTTCAAA 386
QY 932 AGCATGGCTGGATAAATTTGCCCCGGTGTGGGATAATTTAGTCCAAA 978
Db 387 ACGATTATACGATGAATTAATACATAGTAGGGATAGTTTGTATGAAA 433

RESULT 13
US-08-985-335-4/c
Sequence 4, Application US/08985335
Patent No. 6080847
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Yue, Henry
APPLICANT: Lal, Preeti
APPLICANT: Shah, Purvi
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: PROTEINS ASSOCIATED WITH CELL
TITLE OF INVENTION: PROLIFERATION
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Dr.
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/985,335
FILING DATE: Filed Herewith
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0421 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 2082 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:

Db 2664 TTGGCTCGTAAAAATGAAGAAATGTACCTCTTACCTCTTACCTCTCTGAGCCCTTACTCTGTAGAAAAAGAATAC 2605
QY 750 GCAGTGAACAAGATTACACACAACTGGCTTTAAAGATCAAAATGAAATGTTATCAAGTC 807
Db 2604 GGTGTGCCATTCAATTACATCGACATTACTGAAAAATACGATGAATTGGTCGCAATC 2547

Search completed: April 25, 2003, 17:45:05
Job time : 91.5741 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OMnucleic - nucleic search, using sw model

Run on: April 25, 2003, 17:39:49 ; Search time 132.952 Seconds
(without alignments)
16376.941 Million cell updates/sec

Title: US-09-845-416-14_COPY_1000_3000
Perfect score: 2001
Sequence: 1 ggcagttcattgatggagag.....cattgagccaagtgtccgga 2001

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 709820 seqs, 544064369 residues

Total number of hits satisfying chosen parameters: 1410746

Minimum DB seq length: 0
Maximum DB seq length: 5000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_NA.*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq.*
- 2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq.*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq.*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq.*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq.*
- 6: /cgn2_6/ptodata/1/pubpna/PCTUS_PUBCOMB.seq.*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq.*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq.*
- 9: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq.*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq.*
- 11: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq.*
- 12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq.*
- 13: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq.*
- 14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	180	9.0	476	10	US-09-864-761-15766 Sequence 15766, A
2	160.2	8.0	256	10	US-09-864-761-21956 Sequence 21956, A
3	144	7.2	466	10	US-09-864-761-6092 Sequence 6092, Ap
4	141.4	7.1	467	10	US-09-864-761-11083 Sequence 11083, A
5	139	6.9	151	10	US-09-864-761-27715 Sequence 27715, A
6	122	6.1	122	10	US-09-864-761-32272 Sequence 32272, A
7	54.2	2.7	449	9	US-09-918-995-24084 Sequence 24084, A
8	54.2	2.7	2247	10	US-09-960-253-157 Sequence 157, App
9	45.2	2.3	423	10	US-09-864-761-18355 Sequence 18355, A
10	42.4	2.1	2704	10	US-09-205-658-44 Sequence 44, Appl
11	42.4	2.1	2704	10	US-09-844-353A-44 Sequence 44, Appl
12	42.4	2.1	3499	10	US-09-205-658-43 Sequence 43, Appl
13	42.4	2.1	3499	10	US-09-844-353A-43 Sequence 43, Appl
14	41.8	2.1	440	9	US-10-184-644-442 Sequence 442, App
15	41.8	2.1	440	9	US-10-184-634-442 Sequence 442, App
16	40	2.0	425	10	US-09-960-352-4010 Sequence 4010, Ap
17	39.8	2.0	2499	10	US-09-842-552-92 Sequence 92, Appl
18	39.4	2.0	2170	9	US-09-854-133-728 Sequence 728, Appl
19	39.2	2.0	453	10	US-09-861-451A-75 Sequence 75, Appl

C 20	38.4	1.9	590	10	US-09-864-761-8973	Sequence 8973, Ap
C 21	37.4	1.9	1893	9	US-10-227-884-13	Sequence 13, Appl
C 22	37.4	1.9	1893	9	US-10-230-163-13	Sequence 13, Appl
C 23	37.4	1.9	1893	9	US-10-218-631-13	Sequence 13, Appl
C 24	37.4	1.9	1893	9	US-10-230-338-13	Sequence 13, Appl
C 25	37.4	1.9	1893	9	US-10-230-414-13	Sequence 13, Appl
C 26	37.4	1.9	1893	9	US-10-216-159A-13	Sequence 13, Appl
C 27	37.4	1.9	1893	9	US-10-218-849-13	Sequence 13, Appl
C 28	37.4	1.9	1893	9	US-10-227-873-13	Sequence 13, Appl
C 29	37.4	1.9	1893	9	US-10-227-883-13	Sequence 13, Appl
C 30	37.4	1.9	2002	10	US-09-964-824A-264	Sequence 264, App
C 31	37.4	1.9	2082	10	US-09-894-657-4	Sequence 4, Appli
C 32	37.2	1.9	1839	10	US-09-815-242-9459	Sequence 9459, Ap
C 33	37.2	1.9	2160	10	US-09-092-218-1	Sequence 1, Appli
C 34	37.2	1.9	2160	10	US-09-815-242-9181	Sequence 9181, Ap
C 35	37.2	1.9	2915	8	US-08-834-666A-5	Sequence 5, Appli
C 36	37	1.8	1721	9	US-10-043-487-126	Sequence 126, App
C 37	36.8	1.8	2705	10	US-09-925-301-420	Sequence 420, App
C 38	36.4	1.8	454	10	US-09-867-701-2218	Sequence 2218, Ap
C 39	36.4	1.8	4167	9	US-10-091-483-345	Sequence 345, App
C 40	36.4	1.8	4167	9	US-10-079-854-282	Sequence 282, App
C 41	36.4	1.8	4167	10	US-09-764-878-282	Sequence 1145, Ap
C 42	36.4	1.8	4167	10	US-09-764-860-1145	Sequence 345, App
C 43	36.4	1.8	4167	10	US-09-764-846-345	Sequence 212, App
C 44	36	1.8	747	9	US-10-123-155-212	Sequence 637, App
C 45	35.6	1.8	441	9	US-10-025-380-637	

ALIGNMENTS

RESULT 1
US-09-864-761-15766
; Sequence 15766, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aecomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30

; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 15766
; LENGTH: 476
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004468.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.41
US-09-864-761-15766

Query Match 9.0%; Score 180; DB 10; Length 476;
Best Local Similarity 100.0%; Pred. No. 5.2e-43;
Matches 180; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 822 GTTTAAAGCGGATCTAGAAAAGAAAAGCAATCCATGGGCAAACTGTATTCACTCAA 881
|||||
Db 294 GTTTAAAGCGGATCTAGAAAAGAAAAGCAATCCATGGGCAAACTGTATTCACTCAA 353
|||||

QY 882 CAAGATCTTTTCAACTGAAGAATAAGTCAGTGACCCAGAACGCGGAAGCATGGCTG 941
|||||
Db 354 CAAGATCTTTTCAACTGAAGAATAAGTCAGTGACCCAGAACGCGGAAGCATGGCTG 413
|||||

QY 942 GATAACTTTGCCCGGTGGGATAATTTAGTCCAAAACCTTGAAAAAGAGTACAGCACAG 1001
|||||
Db 414 GATAACTTTGCCCGGTGGGATAATTTAGTCCAAAACCTTGAAAAAGAGTACAGCACAG 473
|||||

RESULT 2
US-09-864-761-21956
; Sequence 21956, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663

; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 21956
; LENGTH: 256
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004468.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
; OTHER INFORMATION: SWISSPROT HIT: P11532, EVALUE 2.00e-30
; OTHER INFORMATION: NT HIT: g15032282, EVALUE 1.00e-103
US-09-864-761-21956

Query Match 8.0%; Score 160.2; DB 10; Length 256;
Best Local Similarity 98.2%; Pred. No. 2.6e-37;
Matches 162; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 GGCAGTTCATGTGAGAGAGTGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAGAA 60
|||||
Db 31 GGCAGTTCATGTGAGAGAGTGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAGAA 90
|||||

QY 61 GTATTATCGTGGCTTCTTTCTGCTGAGGACACATTCGAAGCACAAGGAGAGATTTCTAAT 120
|||||
Db 91 GTATTATCGTGGCTTCTTTCTGCTGAGGACACATTCGAAGCACAAGGAGAGATTTCTAAT 150
|||||

QY 121 GATGTGAAGTGGTGAAGACCAAGTTCATCATCTCATGAGGGGTAC 165
|||||
Db 151 GATGTGAAGTGGTGAAGACCAAGTTCATCATCTCATGAGGGTAAAC 195
|||||

RESULT 3
US-09-864-761-6092
; Sequence 6092, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharron G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aeomica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; PRIOR FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359

PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 6092
LENGTH: 466
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC004468.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.8
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.7
US-09-864-761-6092

Query Match 7.2%; Score 144; DB 10; Length 466;
Best Local Similarity 100.0%; Pred. No. 2.6e-32;
Matches 144; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GGCAGTTCATTGATGAGAGTGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAGAA 60
|||||
Db 323 GGCAGTTCATTGATGAGAGTGAAGTAAACCTGGACCGTTATCAAAACAGCTTTAGAAGAA 382
QY 61 GTATTATCGTGGCTTCTTCTGCTGAGGACACATTCGCAAGCACAAAGGAGAGATTCTTAAT 120
|||||
Db 383 GTATTATCGTGGCTTCTTCTGCTGAGGACACATTCGCAAGCACAAAGGAGAGATTCTTAAT 442
QY 121 GATGTGGAAGTGGTGAAAGACCAG 144
|||||
Db 443 GATGTGGAAGTGGTGAAAGACCAG 466

RESULT 4
US-09-864-761-11083
; Sequence 11083, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; GENE EXPRESSION ANALYSIS BY MICROARRAY

FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 11083
LENGTH: 467
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC004468.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.82
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.89
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.84
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.96
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.88
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.81
US-09-864-761-11083

Query Match 7.1%; Score 141.4; DB 10; Length 467;
Best Local Similarity 98.7%; Pred. No. 1.5e-31;
Matches 153; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
QY 339 CAATTTACATAGAGTTTAAATGGATCTCCAGATC-GAAACTGAAAGAGTTGAATGACTG 397
|||
Db 313 CAGTTTACATAGAGTTTAAATGGATCTCCAGATCAGAAACTGAAAGAGTTGAATGACTG 372
QY 398 GCTAACAAAAACAGAGAAGAAACAAAGGAAATGGAGGAGCCCTCTTGGACCTGATCT 457
|||||
Db 373 GCTAACAAAAACAGAGAAGAAACAAAGGAAATGGAGGAGCCCTCTTGGACCTGATCT 432
QY 458 TGAAGACCTAAACCGCCCAAGTACACACATAAGG 492
|||||
Db 433 TGAAGACCTAAACCGCCCAAGTACACACATAAGG 467
RESULT 5


```
US-09-864-761-27715
; Sequence 27715, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 27715
; LENGTH: 151
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004468.1
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.82
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.89
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 0.84
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 0.96
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 0.88
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 0.81
; OTHER INFORMATION: NT HIT: M18533.1, EVALUE 6.00e-80
; OTHER INFORMATION: SWISSPROT HIT: P11532, EVALUE 4.00e-08
US-09-864-761-27715
Query Match 6.9%; Score 139; DB 10; Length 151;
Best Local Similarity 99.3%; Pred. No. 3.7e-31;
Matches 150; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
QY 342 TTTACATAGAGTTTAAATGGATCTCCAGATC-GAAACTGAAAGAGTTGAATGACTGGCT 400
|||||
```

```
Db 1 TTTACATAGAGTTTAAATGGATCTCCAGATCAGAAACTGAAAGAGTTGAATGACTGGCT 60
QY 401 AACAAAAACAGAGAAAGAAACAAAGGAAAAATGGAGGAGAGCCTCTTGGACCTGATCTTGA 460
|||
Db 61 AACAAAAACAGAGAAAGAAACAAAGGAAAAATGGAGGAGAGCCTCTTGGACCTGATCTTGA 120
QY 461 AGACCTAAAAACGCCAAGTAGTACAACAAACATAAG 491
|||
Db 121 AGACCTAAAAACGCCAAGTAGTACAACAAACATAAG 151
RESULT 6
US-09-864-761-32272
; Sequence 32272, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
; FILE REFERENCE: Aemica-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 32272
; LENGTH: 122
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC004468.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 0.41
; OTHER INFORMATION: EST_HUMAN HIT: BE781438.1, EVALUE 1.60e+00
; OTHER INFORMATION: SWISSPROT HIT: P11532, EVALUE 1.00e-17
; OTHER INFORMATION: NT HIT: X06178.1, EVALUE 9.00e-63
```


S-09-864-761-32272
Query Match 6.1%; Score 122; DB 10; Length 122;
Best Local Similarity 100.0%; Pred. No. 3.6e-26;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
y 880 AACAGATCTTCTTCAACACTGAAGAATAAGTCAGTGACCCAGAGACGGAAGCATGGC 939
b 1 AACAGATCTTCTTCAACACTGAAGAATAAGTCAGTGACCCAGAGACGGAAGCATGGC 60
y 940 TGGATAACTTTGCCCGGTGTGGGATAATTTAGTCCAAAAAACTTGAAAAAGAGTACAGCAC 999
b 61 TGGATAACTTTGCCCGGTGTGGGATAATTTAGTCCAAAAAACTTGAAAAAGAGTACAGCAC 120
y 1000 AG 1001
b 121 AG 122

RESULT 7
S-09-918-995-24084
Sequence 24084, Application US/09918995
Publication No. US20030073623A1
GENERAL INFORMATION:
APPLICANT: Hyseq, Inc.
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED
FROM VARIOUS CDNA LIBRARIES
FILE REFERENCE: 20411-756
CURRENT APPLICATION NUMBER: US/09/918,995
CURRENT FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US/09/235,076
PRIOR FILING DATE: 1999-01-20
NUMBER OF SEQ ID NOS: 38054
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 24084
LENGTH: 449
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(449)
OTHER INFORMATION: n = A,T,C or G
S-09-918-995-24084

Query Match 2.7%; Score 54.2; DB 9; Length 449;
Best Local Similarity 51.9%; Pred. No. 1.3e-05;
Matches 122; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
y 1743 CTGAACCTGGCTGCTGAATGTTTATGATACGGGACCAACAGGAGGATCCGTGCTCTGTCT 1802
b 198 CTCAACTTTATGATGCTGTCATATGACAGTGAGGCGGAGGCAAGTTGACGGTATTTTCA 257
y 1803 TTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAAGTACAGATACCTT 1862
b 258 GTTAAAGCTATGTTAGCAACCATGTGTGGTGGAATAATGCTGGACAAATTTAGATATGTT 317
y 1863 TTCAAGCAAGTGGCAAGTTTCAACAGGATTTTGTGACCAAGCGGAGGCTGGCCCTCTCTG 1922
b 318 TTCTCCAGATGTCAGATTCCTCAATGGCTTAATGATATTTAGCAAGTTTGACCAAGTTTCTG 377
y 1923 CATGATTTCTATCCAAATTTCCAAAGACAGTTGGGTGAAGTTGCATCCTTTGGGGCA 1977
b 378 AAGGAAGTTCTGAAGCTCCCAACAGCTGTCTTTGAAGGGCCCATCTTTTGGTTACA 432

RESULT 8
US-09-960-253-157
Sequence 157, Application US/09960253
Patent No. US20020123619A1
GENERAL INFORMATION:
APPLICANT: Benson, Darin R.
APPLICANT: Mohamath, Raodoh
APPLICANT: Lodes, Michael J.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.556
CURRENT APPLICATION NUMBER: US/09/960,253
CURRENT FILING DATE: 2001-09-20
NUMBER OF SEQ ID NOS: 187
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 157
LENGTH: 2247
TYPE: DNA
ORGANISM: Homo sapiens
US-09-960-253-157

Query Match 2.7%; Score 54.2; DB 10; Length 2247;
Best Local Similarity 51.9%; Pred. No. 3.8e-05;
Matches 122; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
y 1743 CTGAACCTGGCTGCTGAATGTTTATGATACGGGACCAACAGGAGGATCCGTGCTCTGTCT 1802
b 517 CTCAACTTTATGATGCTGTCATATGACAGTGAGGCGGAGGCAAGTTTGACGGTATTTTCA 576
y 1803 TTTAAACTGGCATCATTTCCCTGTGTAAAGCACATTTGGAAGACAAAGTACAGATACCTT 1862
b 577 GTTAAAGCTATGTTAGCAACCATGTGTGGTGGAATAATGCTGGACAAATTTAGATATGTT 636
y 1863 TTCAAGCAAGTGGCAAGTTTCAACAGGATTTTGTGACCAAGCGGAGGCTGGCCCTCTCTG 1922
b 637 TTCTCCAGATGTCAGATTCCTCAATGGCTTAATGATATTTAGCAAGTTTGACCAAGTTTCTG 696
y 1923 CATGATTTCTATCCAAATTTCCAAAGACAGTTGGGTGAAGTTGCATCCTTTGGGGCA 1977
b 697 AAGGAAGTTCTGAAGCTCCCAACAGCTGTCTTTGAAGGGCCCATCTTTTGGTTACA 751

RESULT 9
US-09-864-761-18355
Sequence 18355, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharron G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
GENE EXPRESSION ANALYSIS BY MICROARRAY
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/006666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/006662

RESULT 13
US-09-844-353A-43/C
; Sequence 43, Application US/09844353A
; Patent No. US20020037585A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Kimura, Koutarou
; APPLICANT: Patterson, Garth

```

/ GENERAL INFORMATION.
/ APPLICANT: Baker, Kevin P.
/ APPLICANT: Chen, Jian
/ APPLICANT: Desnoyers, Luc
/ APPLICANT: Goddard, Audrey
/ APPLICANT: Godowski, Paul J.
/ APPLICANT: Gurney, Austin L
/ APPLICANT: Pan, James
/ APPLICANT: Smith, Victoria
/ APPLICANT: Watanabe, Colin K
/ APPLICANT: Wood, William I.
/ APPLICANT: Zhang, Zemin
/ TITLE OF INVENTION: SECRETED
/ TITLE OF INVENTION: ACIDS E
/ FILE REFERENCE: P3430R1C227
/ CURRENT APPLICATION NUMBER:
/ CURRENT FILING DATE: 2002-0
/ Prior Application removed -
/ NUMBER OF SEQ ID NOS: 612
/ SEQ ID NO 442
/ LENGTH: 440
/ TYPE: PR1
/ ORGANISM: Homo Sapien
/
/ US-10-184-644-442

```


Search completed: April 25, 2003, 23:58:46
Job time : 148.119 secs

```
Query Match          2.1%; Score 41.8; DB 9; Length 440;
Best Local Similarity 8.7%; Pred. No. 0.06;
Matches 33; Conservative 132; Mismatches 214; Indels 0; Gaps 0;
```

Y 1263 GACCTGAACACCAGATGGAAGCTTCTGCAGGTGGCCGTCGAGGACCGAGTCAGGCAGCTG 1322
|:: :: || : | :: | : |:: :: |:: :: |:: ::

D 429 GDS.WB..TACTYCM.KGKBYSWSCNN...A.CBBNGKBNRBHSY..TTY.NSM.H.C. 370

Y 1323 CATGAAGCCCCACAGGGACTTTGGTGCCAGCATCTCAGCACATTTCTTTCCACGTCTGTCCAG 1382
:: : ::||: : ::||: : : : : : : : : :|:::